

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

<http://go.warwick.ac.uk/wrap/3955>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

**DISCOUNT BAYESIAN MODELS  
AND  
FORECASTING**

**JAMAL RASUL MOHAMMAD AMEEN**

Department of Statistics,  
University of Warwick,  
Coventry CV4 7AL

**DISCOUNT BAYESIAN MODELS  
AND  
FORECASTING**

**JAMAL . R . M . AMEEN**

**PH . D .**

**DEPARTMENT OF STATISTICS**

**UNIVERSITY OF WARWICK**

**MAY 1984**

## **TABLE OF CONTENTS**

	page
<b>1- CHAPTER ONE : INTRODUCTION</b>	
1.1 Status	1
1.2 Outline of the Thesis	3
<b>2- CHAPTER TWO : DISCOUNT WEIGHTED ESTIMATION</b>	
2.1 Introduction	5
2.2 Exponential Weighted Regression	5
2.2.1 The model	5
2.2.2 EWR and time series	7
2.2.3 Some comments on EWR	7
2.3 The simultaneous adaptive forecasting	8
2.4 Discount weighted estimation	10
2.4.1 The model	10
2.4.2 DWE for time series	11
2.5 Applications	13
2.5.1 A simple linear growth model	13
2.5.2 A practical example: The U.S. Air Passenger data set	15
2.6 Summary	19
<b>3- CHAPTER THREE : -YNAMIC LINEAR MODELS</b>	
3.1 Introduction	20
3.2 The DLM's	20
3.3 Relation between DLM's and DWE's	22
3.4 Some limitations and drawbacks	23
3.5 Summary	26
<b>4- CHAPTER FOUR : NORMAL DISCOUNT BAYESIAN MODELS</b>	

4.1 Introduction	27
4.2 Normal Weighted Bayesian Models	28
4.3 Normal Discount Bayesian Models	31
4.3.1 The model	31
4.3.2 Forecasting and updating with NDBM's	33
4.3.3 Coherency	33
4.3.4 Sequential analysis of designed experiments	35
4.4 Other important special cases	36
4.4.1 The modified NDBM	36
4.4.2 Extended NDBM's	38
4.5 Summary	40
<b>5- CHAPTER FIVE : ON-LINE VARIANCE LEARNING</b>	
5.1 Introduction	41
5.2 The Bayesian approach	42
5.3 Non Bayesian methods : A short review	43
5.4 The power law	44
5.5 Summary	47
<b>6- CHAPTER SIX : LIMITING RESULTS</b>	
6.1 Introduction	48
6.2 Similar models and reparameterisation	48
6.3 A common canonical representation	53
6.4 A general limiting theorem	60
6.5 Relations with ARIMA models	62
6.6 Summary	63
<b>7- CHAPTER SEVEN : MULTIPROCESS MODELS WITH CUSUMS</b>	
7.1 Introduction	64
7.2 Historical background and developments	66
7.3 The backward CUSUM	68

7.4 Normal weighted Bayesian multiprocess models	70
7.5 Multiprocess models with CUSUM's	73
7.6 Summary	76
8- CHAPTER EIGHT : APPLICATIONS	
8.1 Introduction	77
8.2 A simulated series	79
8.2.1 Simulation of the data	79
8.2.2 Intervention	80
8.2.3 Multiprocess models - The artificial data	83
8.3 The prescription series	87
8.3.1 The data	87
8.3.2 NDB- multiprocess models : Known observation variance	87
8.3.3 The CUSUM multiprocessor : Known observation variance	90
8.4 The road death series	91
8.4.1 The data	91
8.4.2 The NDB multiprocess model with CUSUM's	91
8.5 Summary	95
9- CHAPTER NINE : DISCUSSION AND FURTHER RESEARCH	
10-APPENDIX	100
11-REFERENCES	102

## ACKNOWLEDGMENTS

I would like to acknowledge my great indebtedness and convey an expression of many thanks to Professor P. J. Harrison for his assistance, guidance, and encouragement throughout the preparation of this work. Thanks also to the members of staff and my fellow students at the Department of Statistics, University of Warwick for many valuable discussions and the Computer Unit for their helpful assistance and facilities.

Finally I would like to thank the University of Sulaimaniyah ( Salahuddin at present) and the Ministry of High Education and Scientific Research-Iraq for the financial support.

To those

I love so much

I owe so much



## SUMMARY

This thesis is concerned with Bayesian forecasting and sequential estimation. The concept of multiple discounting is introduced in order to achieve parametric and conceptual parsimony. In addition, this overcomes many of the drawbacks of the Normal Dynamic Linear Model (DLM) specification which uses a system variance matrix. These drawbacks involve ambiguity and invariance to the scale of independent variables. A class of Normal Discount Bayesian Models (NDBM) is introduced to overcome these difficulties. Facilities for parameter learning and multiprocess modelling are provided. Unlike the DLM's, many limiting results are easily obtained for NDBM's. A general class of Normal Weighted Bayesian Models (NWBM) is introduced. This includes the class of DLM's as a special case. Other important subclasses of Extended and Modified NWBM's are also introduced. These are particularly useful in modelling discontinuities and for systems which operates according to the principle of Management by Exception. A number of illustrative applications are given.

# CHAPTER ONE

## INTRODUCTION

### 1.1. STATUS :

The study of processes that are subject to sequential developments, has occupied scientists for a long time and is currently one of the most active topics in statistics. Indeed, in the majority of real life problems, information arrives sequentially according to some index, often time, and it is desired to detect its plausible pattern and hidden characteristics in order to facilitate control, reduce noise and obtain more reliable estimates and future predictions. The areas of economics, quality control and control engineering are full of such examples. See Whittle (1969), Astrom (1970), Young(1974). In the past, passive procedures ( non Bayesian ) have been used to analyse time series processes. The most popular procedure seem to be through model construction. Models can be classified broadly into two different categories. one of these is called Social Models. Social models provide structures which govern the way the environment behaves. Social or political organisations are members of this class. The other class may be called Scientific Models, These aim to build structures which fit specific environmental characteristics as closely as possible. An important subclass which is the concern of this thesis concerns environments that contain elements of chance. The aim is to build models and measure their adequacy in order to obtain a deeper understanding of the causal mechanism governing the environment. This subclass of Scientific Models is called Statistical Models with mathematics and statistics as its principle tools. Throughout the thesis, models are meant to be Statistical Models unless specified otherwise.

In the classical sense, a time series is a sequential series of observations on a phenomenon which evolves with time. Wold(1954), suggested that a time series process can be decomposed into deterministic components like trend and seasonality with a

random component caused by measurement errors. Before the appearance of computers, among the common short term forecasting procedures, the so called Moving Average criterion was used to fit polynomial functions through least squares. This is reviewed in Kendall, Stuart and Ord (1983). See also Anderson (1977) for further references. With the development of computers, the most widely used models in forecasting during the late 50's were the Exponential Weighted Moving Averages (EWMA) and Holts growth and seasonal model which later developed into the ICI forecasting method, DOUBTS, embodied in the computer package of MULDO and Brown's Exponential Weighted Regression (EWR), Brown(1963). These models are reviewed in Chapter 2 since they stimulated much of the research described in this thesis.

Another well known and widely used class of models is the Autoregressive Integrated Moving Average (ARIMA) models of Box and Jenkins (1970).

Given a series of observations  $\{y_t\}$  and uncorrelated random residuals  $\{a_t\}$ , having a fixed distribution, usually assumed Normal, with zero mean and a constant variance, an ARIMA(p,d,q) is defined in the notation of Box and Jenkins by:

$$(1 + \phi_1 B + \dots + \phi_p B^p)(1 - B)^d y_t = (1 + \theta_1 B + \dots + \theta_q B^q) a_t$$

where B is the backward shift operator,  $B y_t = y_{t-1}$ , and  $\phi_1, \dots, \phi_p ; \theta_1, \dots, \theta_q ; p, q$  and  $d$  are constants whose values belong to a known domain, and are to be estimated from the available data (parameters in a non Bayesian sense).

Despite the existence of a vast amount of literature, these models depend on a large number of unknown constants, that are often difficult to interpret since they do not have natural descriptive meanings. Further, for estimation using the recommended mean square error criterion, a considerable amount of past data is required. Moreover, the resulting models are not robust. They demand stationarity or derived stationarity and make intervention in the form of subjective information difficult. For example, discontinuities and sudden changes can ruin all the estimates.



State Space representations and the works of Kalman and Bucy (1961) and Kalman (1963) have gained considerable grounds regarding fast performance and reduced computer storage problems. However, a natural recipe would have a fully Bayesian representation in which the uncertainty about all the unknown quantities is expressed through probability distributions. The Bayesian Dynamic Linear Models of Harrison and Stevens (1971, 1976) provide such a foundation. This is reviewed in Chapter 3 and a number of limitations and drawbacks are pointed out.

In Bayesian terminology, a time series process is defined to be a parameterised process  $\{Y_t | \theta_t\}$  possessing a complete joint probability distribution for all  $t$ . Initially, there is available prior information ( that is incorporated in the process analysed ) about the parameter vector  $\theta_t$ . This definition is adopted from Chapter 3 and onwards. Vectors are represented by bold phase small letters while capital bold phase letters are used for matrices except the random vector  $Y_t$ .

## 1.2. OUTLINE OF THE THESIS :

In Chapter 2 the one discount factor Exponential Weighted Regression (EWR) method of Brown(1963) and the Simultaneous Adaptive Forecasting of Harrison (1967) are reviewed with some critical comments. The EWR method is then exploited using the discount principle to introduce the general Discount Weighted Estimation ( DWE ) technique. This includes the EWR method, allows different discount factors to be assigned to different model components and provides a preparatory ground for introducing the discount principle into Bayesian Modelling. The method is then applied to the U.S. Air Passengers series and the results are compared with those of DOUBTS and Autoregressive Integrated Moving Average ( ARIMA ) models. The Dynamic Linear Models (DLM's) of Harrison and Stevens(1976) are reviewed in Chapter 3 . Given some initial prior assumptions, for each DWE model, there is a DLM with an identical forecast function. Some limitations and drawbacks of the DLM's are also pointed out.

The discount principle is introduced into Bayesian modelling in Chapter 4 through Normal Weighted Bayesian Models ( NWBM's ). This includes the class of DLM's as a special case. Other important and parsimonious subclasses like Normal Discount Bayesian Models (NDBM's), Modified NDBM's and Extended NDBM's are also introduced. The possibility of including time series models with correlated observations, and some brief comments on the coherency of these models and their relation with the previous models are given. A short outline of the existing on-line estimation of the observation variance is given in Chapter 5 and practical procedures are introduced for variances that have some known pattern or move slowly with time.

Chapter 6 is devoted to reparameterisations and limiting results. Given the eigenstructure of any NDBM, transformations to similar canonical forms are available. A direct procedure is provided for calculating the limiting adaptive factors with no reference to the state precision or covariance matrices. In practice, a limiting state distribution is often quickly reached. This makes such limiting results useful and saves unnecessary computations such as the adaptive vector. Given the adaptive factors, limiting state variance and precision matrices can be calculated independent of each other. Results for some canonical formulations are also given. Limiting NWBM's predictors are compared with those of ARIMA models. This leads to generalisations of some previous results.

In Chapter 7, the principle of Management by Exception and its use in forecasting is sketched. In Bayesian forecasting, the use of multiprocess models had largely replaced the backward Cumulative Sum (CUSUM) of the one step ahead forecasting errors for detecting changes and departures of the process level from specific targets. However, CUSUM's are reintroduced to forecasting systems and operate with multiprocess models which are based on the discount principle. These provide both economical and efficient models called Multiprocess NDB Models with CUSUM's. A number of applications having different characteristics are considered in Chapter 8.

Finally a general discussion is given in Chapter 9. Attention is also paid to further work in progress, and to possible directions for future research.

## CHAPTER TWO

### DISCOUNT WEIGHTED ESTIMATION

#### 2.1. INTRODUCTION:

Operational simplicity and parsimony are among the desirable properties in model constructions. The word 'parsimony' is used here in the sense of Roberts and Harrison (1984). The order of parsimony is the number of unknown constants involved in the model construction. Brown (1963) developed the Exponentially Weighted Regression (EWR), minimising the 'discounted' sum of squares of the one step ahead forecasting errors. As the method depends on one discount factor, it has parsimony of order 1. It will be evident in the coming sections, as is the case in many forecasting methods, that the information content of past observations about the future state of the process decays with its age and this is accomplished using discount factors. The discount concept is a key issue of the thesis and will be exploited in this and the later chapters.

In this chapter, Exponential Weighted Regression is reviewed in Section 2.2 with the emphasis being on time series construction. The DOUBTS method is reviewed in 2.3 and the Discount Weighted Estimation method is introduced in 2.4., Ameen and Harrison (1983 a). This generalisation of EWR uses discount matrices and provides simple recurrence updating formulas. In Section 2.5 a simple linear growth seasonal DWE model is constructed and a practical application is given using the U.S.Air passenger data series. The results are compared with those of DOUBTS and Box-Jenkins.

#### 2.2. EXPONENTIAL WEIGHTED REGRESSION

##### 2.2.1. The Model

One general locally linear representation of a time series process at any time,  $t$ , with future outcomes  $Y_{t+k}$  is :

$$Y_{t+k} = f_{t+k} \theta_{t,k} + \epsilon_{t+k} \quad ; \quad \epsilon_{t+k} \sim [0, V] \quad (2.1)$$

where the components of  $f_{t+k} = [f(1), f(2), \dots, f(n)]_{t+k}$  are independent variables or known functions of time,  $\theta'_{t,k} = [\theta(1), \theta(2), \dots, \theta(n)]_{t,k}$  are unknown with the subscript  $t, k$  indicating that their estimates are based on the data available up to and including time  $t$ , and  $\epsilon_{t+k}$  is a random error term ( $\epsilon_{t+k} \sim [0, V]$  is short hand for the mean of  $\epsilon_{t+k}$  being 0 and the variance  $V$ ).

Usually,  $\theta_{t,k}$  and  $V$  are called the parameters of the model and in a Bayesian sense, they have associated prior distributions. However, in EWR models, these are assumed as constants for the past data  $D_t = \{(y_t, f_t), \dots, (y_1, f_1)\}$ . Given a discount factor  $0 < \beta \leq 1$ ,  $\theta_{t,k} = \theta$  is estimated by  $m_t$  as the value of  $\theta$  that minimises the discounted sum of squares:

$$S_t = \sum_{i=0}^{t-1} \beta^i (y_{t-i} - f_{t-i} \theta)^2 \quad (2.2)$$

Differentiating (2.2) with respect to  $\theta$  at  $\theta = m_t$ , and equating the result to 0,

$$\sum_{i=0}^{t-1} \beta^i f'_{t-i} (y_{t-i} - f_{t-i} m_t) = 0 \quad (2.3)$$

Now, define

$$Q_t = \sum_{i=0}^{t-1} \beta^i f'_{t-i} f_{t-i} = \beta Q_{t-1} + f'_t f_t \quad (2.4)$$

$$k_t = \sum_{i=0}^{t-1} \beta^i f'_{t-i} y_{t-i} = \beta k_{t-1} + f'_t y_t \quad (2.5)$$

Assuming that  $Q_t^{-1}$  is the generalised inverse of  $Q_t$ , it can be seen from (2.3), that

$$m_t = Q_t^{-1} k_t$$

This, with (2.4) and (2.5), gives the following relationship on  $m_t$

$$m_t = m_{t-1} + a_t e_t$$

where  $a_t = Q_t^{-1} f'_t$  and  $e_t = y_t - f_t m_{t-1}$ . The  $k$  steps ahead point forecast is given by



$$f_{t+k} = f_t$$

### 2.2.2. EWR and Time Series

In time series processes, the form of the forecast function can often be specified up to a reasonable degree of approximation. General polynomial predictors can be constructed through specifications of the design vector,  $f_{t+k}$ . A simple and efficient way, as presented by Brown (1963), is to define  $f_{t+k} = f_t G^k$ , where  $G$  is a non singular matrix with dimension  $n$ . Therefore, using the notations and the criterion of Section 2.2.1, with  $f_t = f$  being independent of time, the alternative forms of (2.4) and (2.5) are :

$$Q_t = \beta G'^{-1} Q_{t-1} G^{-1} - f' f \quad (2.6)$$

and

$$h_t = \beta G^{-1} h_{t-1} + f' y_t \quad (2.7)$$

The current estimate of  $\theta$  at time  $t$ , is then given by

$$m_t = G m_{t-1} + a_t e_t$$

where

$$e_t = y_t - f G m_{t-1} \quad \text{and} \quad a_t = Q_t^{-1} f'$$

### 2.2.3. Some Comments on EWR

In order to get some insight into the terms and equations obtained in Sections 2.2.1 and 2.2.2, consider the minimisation of (2.2) again. Note that the same estimates of  $\theta$  can be obtained by maximising  $L = \exp\{-S_t/(2V)\}$  for  $\theta$ . Given that  $\epsilon_{t+k}$  is a Normal random variable,  $L$  can be called the 'likelihood function' of  $\theta$  at time  $t$ .  $Q_t$ , in (2.4) and (2.6) is proportional to the so called Fisher's Information matrix about  $m_t$  ( minus the second derivative of  $L$  with respect to  $\theta$  at  $\theta = m_t$  ), or the 'precision' matrix in a Bayesian sense. The proportionality constant is  $V$ . Moreover, in (2.6),  $Q_t$  is decomposed into  $f' f$ , the information content from the most recent observation and  $\beta G'^{-1} Q_{t-1} G^{-1}$ , the information contributed from the past data, discounted by  $\beta$ . This,



together with the convergence of (2.2), restricts the values of  $\beta$  to the range,  $0 < \beta < 1$ . Thus the role of the 'discount factor', describes the rate at which the information about model parameters changes with time. Moreover, given  $\lambda_1, \lambda_2, \dots, \lambda_n$  as distinct eigenvalues of  $G$ , the convergence of (2.6) requires that  $0 < |\beta/\lambda_i|^2 < 1$ . This can be seen on rewriting (2.6) as

$$Q_t = \beta^t G^{t-1} Q_0 G^{-t} + \sum_{i=0}^{t-1} \beta^i G^{t-i-1} f' f G^{-i}$$

Combining the restrictions on  $\beta$ , we have  $0 < \beta < \min\{1, |\lambda_1|^{-2}, |\lambda_2|^{-2}, \dots, |\lambda_n|^{-2}\}$ . The convergence of the adaptive vector  $a_t$  follows from the convergence of  $Q_t$ .

### 2.3. THE SIMULTANEOUS ADAPTIVE FORECASTING:

Consider a time series process that can be decomposed into three different components of trend, seasonality, and random variation. Suppose that the seasonal component changes relatively very slowly, so that the greater percentage of the predictive variation is attributable to changes in trend and random variation (the data analysed at the end of this chapter, is of this type). EWR assumes that the loss of information with age, occurs at the same rate for both the trend and seasonal components, whereas we know that the information on the seasonal component is more durable, and hence, more appropriately discounted using a much higher discount factor than that which is appropriate for the trend component. This led Harrison (1967) to propose an alternative procedure to EWR which considered a simple multiplicative linear growth and seasonal model of parsimony 2. That work led to the development of the forecasting method DOUBTS or, Simultaneous Adaptive Forecasting, which is the basis of the I.C.I. short term forecasting computer package MULDO. Harrison and Scott (1965). Whittle (1965) examined the method. The following is a short review of DOUBTS, with some comments. The  $k$ -steps predictor  $F_t(k)$  is

$$F_t(k) = (m_t + kb_t)S_t(k)$$

where

$$m_t = m_{t-1} + b_{t-1} + (1 - \beta_1^2) e_t$$

$$b_t = b_{t-1} - (1 - \beta_1)^2 e_t$$

$$e_t = y_t - F_{t-1}(1)$$

and  $\beta_1$  is the trend discount factor.

The seasonal component for  $k$  periods ahead, is given by

$$S_t(k) = 1 + \sum_{i=1}^n \{a_i(t) \cos(H_i x_k) + b_i(t) \sin(H_i x_k)\},$$

where there are  $n$  significant harmonics, with  $H_i$  taking the appropriate integer values in

the range 1 to  $T/2$ , and  $a_i(t)$ , and  $b_i(t)$  are the harmonic coefficients at time  $t$ .  $x_k = \frac{2\pi k}{T}$ ,

where  $T$  is the length of the seasonal period.

Given

$$\alpha'_t = [a_1, b_1, \dots, a_n, b_n]_t$$

$$G = \text{diag}\{G_1, G_2, \dots, G_n\}$$

$$G_k = \begin{bmatrix} \cos(x_k) & \sin(x_k) \\ -\sin(x_k) & \cos(x_k) \end{bmatrix}$$

Then

$$\alpha_t = G \alpha_{t-1} + \alpha e'_t$$

$$e'_t = (y_t/m_t) - S_{t-1}(k)$$

where  $\alpha$  is Brown's adaptive constant vector, whose elements are functions of the seasonal discount factor  $\beta_2$ . More details on  $\alpha$  can be found in Harrison and Scott (1965).

Although it is not intended here to proceed with the generalisation of this method, by the end of this chapter it will be evident that higher degree and parsimonious

polynomials with more economical but still efficient seasonal components can be accommodated. However, like its predecessors, the method is limited and suffers from both theoretical and logical justifications. It is purely a point estimator. Unlike Holts seasonal forecasting method, the seasonal effects are included in the trend updating equations while the trend contribution  $m_t$  is removed in updating the seasonal components.

Other means of constructing adaptive vectors for sequential estimation purposes are used through stochastic approximation, Gelb (1974). These provide estimates that are not necessarily optimal in any statistical sense. They possess some desirable, well defined convergence properties irrespective of the parameter uncertainties. See also Maybeck (1982).

## 2.4. DISCOUNT WEIGHTED ESTIMATION :

In this section, the idea of 'discounting', as discussed in Section 2.3, is generalised to 'multiple discounting'. This provides a new class of models called Discount Weighted Estimation (DWE), using different discount factors for different model components.

### 2.4.1. The Model

Let a time series process be represented by (2.1),  $\epsilon_{t+k}$  ( $k > 0$ ) be independent of the data  $D_t = \{(y_t, f_t), D_{t-1}\}$ , and given  $D_t$ ,  $\theta_{t,k} = \theta$ , be estimated by  $m_t$  at time  $t$ .

### DEFINITION

A DWE model is given by

$$E[Y_{t+k} | D_t, f_{t+k}] = f_{t+k} m_t,$$

where :

$$m_t = m_{t-1} + a_t e_t \quad (2.8)$$

$$a_t = Q_t^{-1} f'_t \quad (2.9)$$

$$e_t = y_t - f_t' m_{t-1} \quad (2.10)$$

$$Q_t = B' Q_{t-1} B + f_t' f_t, \quad (2.11)$$

$$B = \text{diag}\{\beta_1, \beta_2, \dots, \beta_n\} ; 0 < \beta_i < 1, \quad i = 1, 2, \dots, n \quad (2.12)$$

The EWR model is retained when  $B = \beta I$  where  $I$  is an identity matrix of order  $n$ . Notice that only  $Q_t^{-1}$  and not  $Q_t$  needs to be calculated to obtain  $m_t$ . Although the inversion method for obtaining  $Q_t^{-1}$  has been around for a long time Henderson et al(1959) and later Lindley and Smith(1972), even now it does not seem to be generally appreciated by practitioners even in EWR case. A more attractive recursion which avoids matrix inversions and their associated problems is to replace (2.9) - (2.11), by :-

$$Q_t^{-1} = (I - a_t f_t) R_t \quad (2.13)$$

$$R_t = B^{-1} Q_{t-1}^{-1} B^{-1} \quad (2.14)$$

$$a_t = R_t f_t' (1 + f_t R_t f_t')^{-1} \quad (2.15)$$

It can be seen from (2.11), that any initial value for  $Q_0$ , and hence  $m_0$ , will be dominated after a small number ( around  $n$ , the dimension of  $\theta$  ) of iterations. In cases of ignorance the initial default settings  $Q_0^{-1} = \alpha I$  and  $m_0$ , where  $\alpha$  is a large number,  $10^5$  ( say ), are usually adequate for operation. However in most cases, there is at least a vague impression of the size of the elements of  $\theta$  which will give a better value of  $m_0$  so that  $f_1 m_0$  is close to  $y_1$ . From 2.2.3, we know that  $Q_0 = V C_0^{-1}$ , where,  $C_0$  represents Fisher's Information matrix about  $\theta$  at time  $t=0$ . Then  $Q_0^{-1}$  can be set by assuming an upper limit for  $V$ , the variance of  $\epsilon$ ; choosing a liberal marginal value  $c_i$  for  $\theta_i$  and setting  $Q^{-1} = \text{diag}\{c_1, c_2, \dots, c_n\} / V$ . These ideas are illustrated by an example in 2.5.

#### 2.4.2. DWE for Time Series

The principle of superposition states that any linear combination of linear models is a linear model. Model builders often use this in reverse, decomposing a linear model and

extending the principle to statistical models using the fact that a normal random vector may be decomposed into a set of components of normal random vectors. The major point is that the component models can be built separately and then combined linearly to obtain a complete model. Hence in practice, given a time series for which  $f_{t+k} = f_t G^k$  for all  $t, k > 0$ , often  $G$  is decomposed into  $r$  components and written  $G = \text{diag}\{G_1, G_2, \dots, G_r\}$ , where  $G$  is non singular. The case of special interest will be covered by assuming that the  $n_i$  square matrix  $G_i$  has a single associated discount factor  $\beta_i$ , and  $\sum_{i=1}^r n_i = n$ .

# DEFINITION

The method of DWE, for time series, is given by the forecast function

$$E[Y_{t+k} | D_t] = f_{t+k} m_t = f G^k m_t,$$

where  $m_t$  is recursively calculated by

$$m_t = G m_{t-1} + a_t e_t$$

$$e_t = y_t - f G m_{t-1}$$

$$a_t = Q_t^{-1} f' = R_t f' (f R_t f' + 1)^{-1},$$

$$Q_t^{-1} = (I - a_t f) R_t \tag{2.16}$$

$$R_t = B^{-1} G Q_{t-1}^{-1} G' B^{-1}$$

and  $B = \text{diag}\{\beta_1 I_1, \beta_2 I_2, \dots, \beta_r I_r\}$  where  $0 < \beta_j < 1$ , and  $I_j$  is an identity matrix of order  $n_j$ ;  $j = 1, 2, \dots, r$ .

# THEOREM 2.1.

For the DWE method defined above, if  $\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,n_i}$   $i = 1, 2, \dots, r$ , are non zero eigenvalues of  $G_i$  and  $\lim_{t \rightarrow \infty} Q_t = Q$  exists for all bounded  $Q_0$ , then



$$0 < \beta_i < \min\{1, |\lambda_{i,1}|^2, |\lambda_{i,2}|^2, \dots, |\lambda_{i,n_i}|^2\}$$

## PROOF

Using (2.16), we have,

$$\begin{aligned} Q_t &= B' G'^{-1} Q_{t-1} G^{-1} B' + f' f \\ &= B^{t-2} G'^{-t} Q_0 G^{-t} B^{t-2} - \sum_{k=0}^{t-1} B^{k-2} G'^{-k} f' f G^{-k} B^{k-2}, \end{aligned} \quad (2.17)$$

since  $B^2$  and  $G^{-1}$  commute.

The convergence of (2.17), gives that

$$0 < |\beta_i / (\lambda_{i,j})^2| < 1 \quad ; i=1,2,\dots,r \quad ; j=1,2,\dots,n_i$$

The result is obtained by combining this with the conditions  $0 < \beta_i < 1$ .

Some modellers have suggested to move beyond these assumptions in order to increase models adaptivity, Muth (1981). Clearly such models produce highly unreliable and statistically unsounded forecasts. A proper way of introducing temporal adaptivity is dealt with later through discounting the prior information.

Under the above assumptions, the recursive formulas converges considerably fast to a limiting form. Apart from computational benefits, as will be seen later, these provide limiting justifications of many commonly used forecasting structures in literature. It also uncovers in spirit, the partial success achieved by some classical models like the ARIMA models.

## 2.5. APPLICATIONS:

### 2.5.1. A simple linear growth seasonal model

Using the principle of superposition for normal random variables, suppose that a time series model can be constructed using a linear combination of a linear growth,

seasonal and random components.

The linear growth model may be described by the pair  $\{f_1, G_1\} = \{[1, 0], \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\}$ . This is evident since  $f_1 G_1^k = [1, k]$ . Then if  $m_t$  and  $b_t$  are the present estimates of the level and growth rate, the forecast function of this component is  $f_1 G_1^k m_t = m_t + kb_t$ , which is the familiar Holt-Winters linear growth function.

Any additive seasonal pattern  $S(1) \dots S(T)$ , for which  $n$  is the integer part of  $(T+1)/2$  and such that  $\sum_{j=1}^n S(j) = 0$ , can be expressed in terms of harmonic functions as

$$S(k) = \sum_{i=1}^n (a_i \cos(kw_i) - b_i \sin(kw_i)) \quad (2.18)$$

where  $w = 2\pi/T$ .

Equation (2.18), can be represented equivalently as

$$S(k) = \sum_{i=1}^n \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \cos iw & \sin iw \\ -\sin iw & \cos iw \end{bmatrix}^k \begin{bmatrix} a_i \\ b_i \end{bmatrix}$$

An alternative seasonal component model which gives an identical performance to that previously discussed is :-

$$f_2 = [f_{2,1}, f_{2,2}, \dots, f_{2,n}]$$

$$G_2 = \text{diag}\{G_{2,1}, G_{2,2}, \dots, G_{2,n}\}$$

where  $f_{2,k} = [1, 0]$  and  $G_{2,k} = \begin{bmatrix} \cos(kw) & \sin(kw) \\ -\sin(kw) & \cos(kw) \end{bmatrix}$ ;  $k=1,2,\dots,n$ . The practical benefit of this harmonic representation occurs when there exists an economical seasonal representation in terms of a limited number of harmonics. For example, Box and Jenkins (1970) examined the mean monthly temperature for Central England in 1964 and demonstrated that over 96 % of the variation can be described by a single first harmonic; the rest of the variation being well described as random. In this case the seasonal pattern is captured by two unknowns rather than eleven as in the full monthly seasonal

description.

In applying DWE it is generally advisable to associate a discount factor  $\beta_1$  with the linear growth component but have a higher discount factor  $\beta_2$  for the seasonal description. This is due to the fact that often the seasonal pattern is more stable than the trend.

The full linear growth seasonal model is then

$$\{f_1, f_2; \text{diag}\{G_1, G_2\}; \text{diag}\{\beta_1 I_2, \beta_2 I_{2n}\}\}$$

where  $I_k$ ,  $k=2, 2n$ , is the identity matrix of dimension  $k$ .

### 2.5.2. A practical Example: The U.S. Air Passenger Data Series

For comparison with other methods the ten years monthly data from 1951 to 1960 is analyzed. The data is given in Brown(1963) and Box and Jenkins (1970). The series is a favourite with analysts since it has strong trend and seasonal components with very little randomness. However, it is not a strong test of a forecasting method. Harrison (1965) showed that the EWR method proposed by Brown cannot achieve a Mean Absolute Deviation (MAD) of less than 3% since it insists upon using a single inadequate discount factor in a case in which the trend and seasonal components require significantly different discount factors. He stated that if, on this data, a method cannot achieve a MAD of less than 3%, then that method can be regarded as suspect. Harrison analyzed the data using the DOUBTS method described in 2.3.

In this section the DWE model  $\{f, G, B\}$  is applied to the logarithms of the data using:-

$$f = [1, 0, 1, 0, \dots, 1, 0], \quad G = \text{diag}\{G_1, G_2, \dots, G_6\}$$

$G_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  for the trend and  $G_{k+1} = \begin{bmatrix} \cos(kw) & \sin(kw) \\ -\sin(kw) & \cos(kw) \end{bmatrix}$  for  $k=1, 2, \dots, 5$  as representing the  $k^{\text{th}}$  harmonic description of the seasonal pattern with  $w = \pi/6$ . The point forecasts were obtained as the exponential of the DWE results.



A pair of discount factors was used with  $\beta_1$  for the trend ( relating to  $G_1$  ) and  $\beta_2$  for the seasonal block.

Initially the prior specifications was

$$\{m_0, Q_0^{-1}\} = \left\{ \begin{bmatrix} 5 \\ .1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & .001 & 0 \\ 0 & 0 & 0.02I \end{bmatrix} \right\}$$

which corresponds to a specification of no seasonal pattern !. a level lying within a 95% interval [ 80 ; 280 ] and a monthly growth of between 4% and 18% per month. Hence this is a very weak prior although it does not assume complete ignorance. Fig.1 presents the one-step-ahead point predictions with the observations.

For comparison, the one step ahead forecast errors over the last six years were obtained and a DWE performance of 2.3% MAD achieved.

Another well known analysis of the data is given in the book of Box and Jenkins. Writing  $x_t$  as the logarithm of the  $t^{th}$  observation and  $a_t$  as the corresponding one step ahead errors, their predictions are obtained using the difference equation:-

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + a_t - \theta a_{t-1} - \psi a_{t-12} + \theta\psi a_{t-13}$$

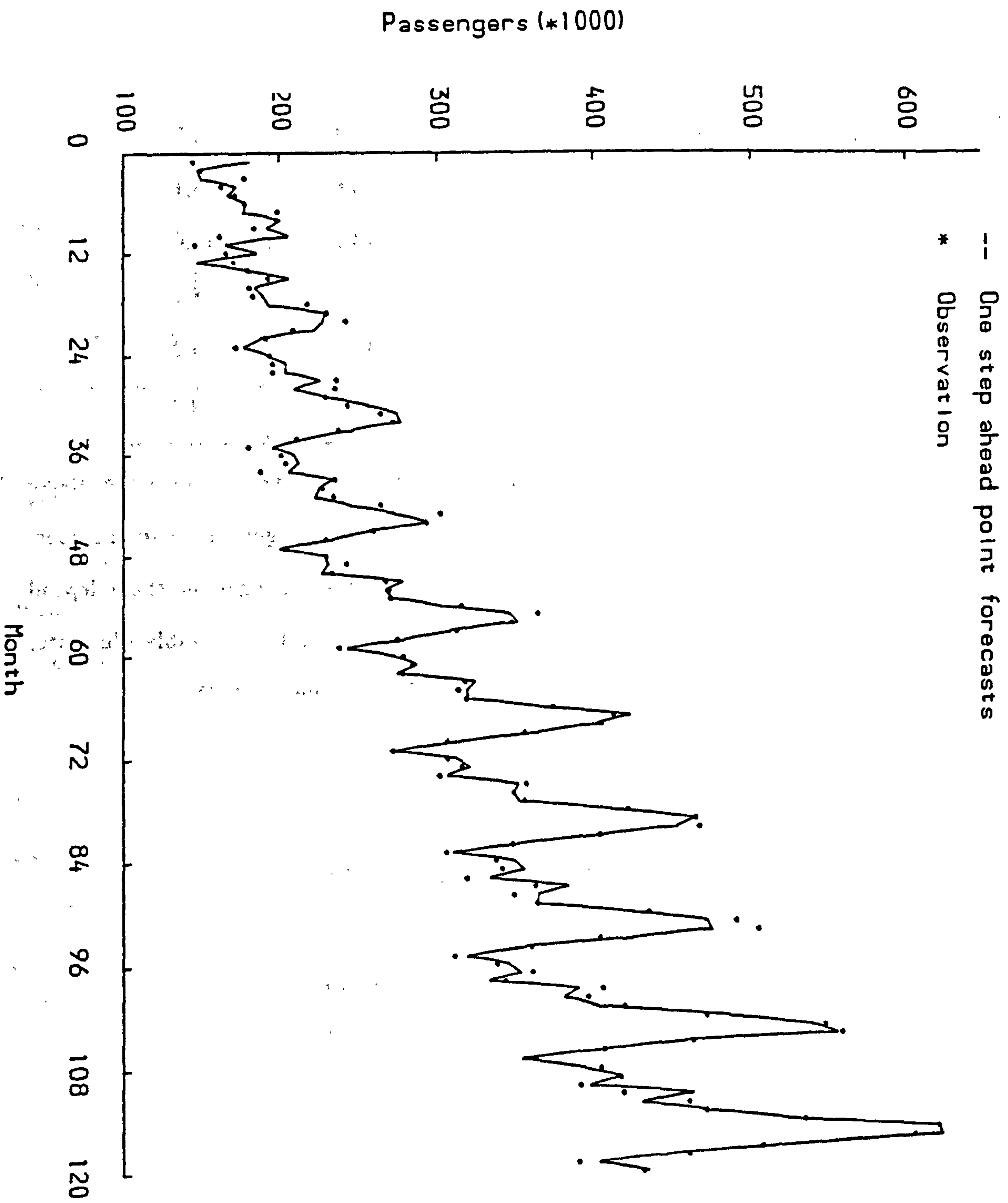
where the mean square error is minimised when  $\theta = .4$  and  $\psi = .61$ . This method is also of parametric parsimony 2 and the following table indicates the comparability of the performance with that of DOUBTS and that of DWE with the same discount pair (0.84 , 0.93) as described in Harrison (1965) and with the discount pair (0.76 , 0.91) which reduces the MAD of the errors.

The Mean Absolute Forecast Errors For The Year 1955-1960				
Year	DOUBTS (.84,.93)	DWE (.84,.93)	DWE (.76,.91)	B & J (.4,.61)
1955	9.4	7.7	7.0	8.0
1956	4.3	5.4	5.4	4.5
1957	5.5	5.5	5.6	6.1
1958	15.2	14.7	13.7	14.0
1959	11.8	11.5	9.8	8.7
1960	11.0	11.5	12.1	14.2
MEAN	9.4	9.4	8.9	9.3

Clearly, in this example, no significant difference is observed between the above results. However, DWE has the properties of being more general, parsimonious, intervention can easily be accommodated in the phase of sudden changes and these depend on a small number of easily assessed discount factors. The following table illustrates models sensitivity for different selection of discount pairs in terms of MAD.

$\beta_2$			
$\beta_1$	0.8	0.9	1.0
0.6	9.71	9.39	15.36
0.7	9.45	9.08	15.3
0.8	10.41	9.12	14.15
0.9	16.56	11.91	13.77
1.0	32.58	22.59	16.44

FIG. 1. U.S. AIR PASSENGER SERIES 1951-1960



## **2.6. SUMMARY:**

The methods of EWR and DOUBTS are reviewed and some general comments , drawbacks and limitations are pointed out. The estimation procedure of DWE is introduced as a fruitful extension of EWR to provide and prepare solid grounds for introducing the discount concept into the Bayesian Forecasting. The method is applied to the U.S.air passenger data set and the results are encouraging when compared with previous existing ones.

## CHAPTER THREE

### DYNAMIC LINEAR MODELS

#### 3.1. INTRODUCTION :

One of the main contributions to Bayesian Statistics, both in theory and applications, is Bayesian forecasting. This provides a natural way of combining experts intelligence with the information provided by the data. The DLM's of Harrison and Stevens (1971, 1976) provide such means. The method also gives limiting justifications for many well known classical forecasting procedures, Brown(1963), Holt(1957), Winters(1960), and Box and Jenkins(1970). State space representations can be found in Priestley (1980). In particular, an extensive amount of literature is available on engineering applications of the Kalman Filter, Kalman(1963). Bayesian Statistics has widened the understanding of random phenomenon and introduced the facilities of on-line variance learning, intervention , multiprocess modelling and has relaxed the assumption of stationarity. The method provides forecast distributions rather than point estimates.

In this chapter, DLM's are reviewed in 3.2 and relationships with DWE methods are discussed in 3.3. The DLM recursive formulas in the parameter estimation are attractive for the ease of elaboration and reduce considerably the computer storage problem. However, the method is not free from drawbacks. The specification of the system matrix  $W$  has caused problems in practice. This, together with other difficulties, is discussed in 3.4. A short summary of the contents of the chapter is given in 3.5

#### 3.2. THE DLM's :

The class of DLM's as defined by Harrison and Stevens (1976) constitute quadruples  $\{F, G, V, W\}_t$  with proper dimensionality. A particular parametrised process  $\{Y_t | \theta_t\}$  can be modelled using this class of models if the following linear relations hold:

$$i) \quad Y_t = F_t \theta_t + v_t \quad ; \quad v_t \sim N[0; V_t] \quad (3.1)$$

$$ii) \quad \theta_t = G_t \theta_{t-1} + w_t \quad ; \quad w_t \sim N[0; W_t] \quad (3.2)$$

The first of these equations is called the observation equation which combines the observable vector  $Y_t$  to an unobservable state parameter vector and an additive error structure which is assumed to be Normally distributed with mean  $0$  and variance  $V_t$ . The second equation describes the evolution of the state vector with time. Unless otherwise stated, the random vectors  $v_t$  and  $w_t$  are assumed to be uncorrelated with known variance matrices  $V_t$  and  $W_t$  respectively.

Given an initial prior  $(\theta_0 | D_0) \sim N[m_0; C_0]$ , using Bayes theorem with  $D_t = \{y_t, D_{t-1}\}$ , it follows that :

$$(Y_t | D_{t-1}) \sim N[\hat{y}_t; \hat{Y}_t] \quad (3.3)$$

$$(\theta_t | D_t) \sim N[m_t; C_t] \quad (3.4)$$

where:

$$\hat{y}_t = F_t G_t m_{t-1} \quad ; \quad \hat{Y}_t = F_t R_t F'_t + V_t \quad (3.5)$$

$$m_t = G_t m_{t-1} + A_t e_t \quad (3.6)$$

$$C_t = (I - A_t F_t) R_t \quad (3.7)$$

$$R_t = G_t C_{t-1} G'_t + W_t \quad (3.8)$$

$$A_t = R_t F'_t \hat{Y}_t^{-1} = C_t F'_t V_t^{-1} \quad (3.9)$$

$$e_t = y_t - \hat{y}_t \quad (3.10)$$

When  $\{F, G, V, W\}$  are all known and are not dependent on time, then the DLM is called a constant DLM

### 3.3. RELATION BETWEEN DLM's AND DWE's :

The updating recurrence relations (3.5)-(3.10), suggests a connection between estimation using DWE and the estimation using DLM's. In order to establish this relationship, we first give the following

#### DEFINITION

For any DWE  $\{f, G, B\}_t$  with initial setting  $(m_0; Q_0)$ , where  $Q_0$  is nonnegative definite, a corresponding DLM is given by  $\{f, G, V, W\}_t$  where  $W_t = (H_t Q_{t-1}^{-1} H'_t - G_t Q_{t-1}^{-1} G'_t) V_t$  is nonnegative definite and  $H_t = B_t^{-1} G_t$ .

In case of  $Q_{t-1}$  being singular,  $Q_{t-1}^{-1}$  represents a generalised inverse of  $Q_{t-1}$ .

#### THEOREM 3.1.

For the DWE  $\{f, G, B\}_t$ , the corresponding DLM produce a forecast function identical to that obtained by DWE. Further,  $(\theta_t | D_t) \sim N[m_t; C_t]$  where  $m_t$  is the DWE estimate and  $C_t = Q_t^{-1} V_t$ .

#### PROOF:

From the initial settings, the theorem is true for  $t=1$ . Using induction, suppose true for  $\{t \text{ minus } 1\}$ . From the DLM results, we have:

i)

$$R_t = G_t Q_{t-1}^{-1} G'_t V_t + W_t = H_t Q_{t-1}^{-1} H'_t V_t$$

and since

$$\begin{aligned} C_t^{-1} &= R_t^{-1} + f'_t f_t V_t^{-1} = [H'_t Q_{t-1}^{-1} H_t + f'_t f_t] / V_t \\ &= Q_t / V_t \end{aligned}$$

we have

$$C_t = Q_t^{-1} V_t$$



ii)

$$E[\theta_t | D_t] = G_t m_{t-1} + a_t e_t = m_t \quad \text{the DWE estimate, since}$$

$$a_t = G_t f'_t / V_t = Q_t^{-1} f'_t, \quad \text{the DWE } a_t$$

iii) The forecast function is

$$E[Y_{t+k} | D_t] = f_{t+k} \prod_{i=1}^k G_{t+i} m_i \quad \text{as for the DWE.}$$

### COROLLARY 3.1.1.

For  $t \geq 0$ , the DWE  $\{f_t, G_t, V, I\}$  gives a forecast function identical to that of the DLM  $\{f_t, G_t, V, 0\}$ .

### PROOF

Obvious since from the definition  $W_t = 0$  for all  $t$ .

In DLM terms the above setting for  $W_t$  is unusual in its dependence upon  $C_{t-1}$ , the uncertainty of the observer concerning  $\theta_t$  given  $D_{t-1}$ . The concept that the observers view of the future development of the process depends upon his current information is also adopted in Entropy Forecasting, Souza(1978) and in Smith (1979).

### 3.4. SOME LIMITATIONS AND DRAWBACKS :

Time series processes are often best described reasonably using parametric statistical models. In this case, efficient model intervention can be performed in various stages of the analysis. Within the Bayesian framework, DLM's are often used for this purpose. However, the latter require experience in the representation of innovations using Normal probability distributions. The specification of the associated system variance matrices has proved a major obstacle. Practical problems arise because of the non uniqueness of  $V$  and  $W$  and because the lack of familiarity with such matrices causes application difficulties and lead practitioners to other methods. Even experienced people find that



they have little natural quantitative feel for the elements of these matrices. Their ambiguity arise since there exists an uncountable number of time shift reparametrisations which have identical forecast distributions. For example:

The constant Normal DLM

$$Y_t = \theta_t + v_t$$

$$\theta_t = \lambda \theta_{t-1} + w_t$$

with  $\begin{bmatrix} v_t \\ w_t \end{bmatrix} \sim N(\mathbf{0}; \mathbf{U} = \begin{bmatrix} V - \alpha S & \alpha S \\ \alpha S & W - \alpha(1 - \lambda^2)S \end{bmatrix})$ , can be represented as

$Y_t - \lambda Y_{t-1} = v_t - \lambda v_{t-1} + w_t$ . This is a stationary process provided that  $|\lambda| < 1$ , and without loss of generality, for an infinite series, it can be written equivalently as

$$Y_t = v_t + \sum_{i=0}^{\infty} \lambda^i w_{t-i}$$

so that

$$\text{Var}(Y_t) = V + W/(1 - \lambda^2),$$

and

$$\text{Cov}(Y_t, Y_{t+k}) = \lambda^k W/(1 - \lambda^2).$$

Provided that  $\mathbf{U}$  is a covariance matrix, the joint distribution of  $Y_t, Y_{t+1}, Y_{t+2}, \dots, Y_{t+k}$  does not depend on  $\alpha$ . i.e; for infinitely many values of the variances of the  $v$ 's and  $w$ 's, the same forecast distribution is obtained. This generalises easily to higher dimensional DLM's.

Attempts are made to estimate  $V$ ,  $W$  and  $C$  using sample autocovariances, see Lee(1980), however, the ambiguity in these systems is always evident unless further constraints are added!. The system error variance is also not invariant to the scale on which the independent variables are measured. To overcome these difficulties, Ameen and Harrison (1983 b) have replaced this system matrix by a discount matrix. The procedure is both easy to understand and simple to elaborate and this is the concern of

the coming chapters.

### **3.5. SUMMARY :**

The class of DLM's is represented in 3.2 and its relation with DWE estimates is given in 3.3 . It is shown that given a DWE model, there exists a DLM having the same forecast function. Limitations and drawbacks of DLM s are discussed in 3.4.

## CHAPTER FOUR

### NORMAL DISCOUNT BAYESIAN MODELS

#### 4.1. INTRODUCTION:

Two desirable properties of applied mathematical models are ease of application and conceptual parsimony. Hence the attraction of discount factors in methods of sequential estimation. Chapter 2 was concerned with the method of DWE which generalises the estimation method of Exponentially Weighted Regression promoted by Brown(1963). In the simplest situation a single discount factor  $\beta$  describes the rate at which information is lost with time so that if the current information is now worth  $M$  units, then its worth with respect to a period  $k$  steps ahead is  $\beta^k M$  units. However if a system has numerous characteristics then the discount factor associated with particular components may be required to take different values. The DWE method provides a means of doing this but it is strictly a point estimation method.

The Bayesian approach to statistics is a logical and profound method. In forecasting it provides information as probability distributions (support (likelihood) functions follow consequently). These are essential to decision makers. The major objective of this work is to provide Bayesian Forecasting methods founded upon the discount concept. This concept has been applied in the ICI forecasting package MULDO as described in 2.3, Harrison(1965) and Harrison and Scott(1965), and the ICI Multivariate Hierarchical forecasting package, Harrison, Leonard and Gazard(1977). The former is a particular method which does not easily generalise, Whittle(1965). The latter and other applications have been based upon Constant Dynamic Linear Models (DLM's) which have limiting forecast functions equivalent to those derived using EWR, Godolphin and Harrison (1975), Harrison and Akram(1983). The use of such models has involved practitioners specifying a system variance matrix  $W$  which has elements that are

proportional to functions of a discount factor. They are thus indirect applications of discounting. This chapter is concerned with a class of Normal Discount Bayesian Models ( NDBM's ) which eliminates the system variance matrix  $W$ . Instead a discount matrix is introduced which associates possibly different discount factors with different model components. Such a discount factor converts the component's posterior precision  $P_{t-1}$  at time  $t-1$  to a prior precision  $P_t^* = \beta P_{t-1}$  for time  $t$ . The term precision is used in its Bayesian sense but may also be thought of as a Fisherian measure of information.

The use of the discount matrix overcomes the major disadvantages of the system variance  $W$ , since ambiguity is removed. the discount matrix is invariant ( up to a linear transformation ) to the scale on which both the independent and dependent variables are measured , and the methods are easily applied. Because of conceptual simplicity and ease of operation it is anticipated that the NDBM approach will find many applications in dynamic regression, forecasting , time series analysis , the detection of changes in process behaviour, quality control and in general statistical modelling where the observations are performed sequentially or ordered according to some index.

In this chapter Normal Weighted Bayesian Models (NWBm) are introduced in 4.2 and their relations with DLM's is pointed out. Particular emphasis is given to a subclass of models called NDBM's and the possibility of retaining the model coherency is discussed in 4.3. Other practically important subclasses of models like the Modified NDBM's and the Extended NDBM's are discussed in 4.4. These extend the capability of the models in cases of sudden changes in the process behaviour and in dealing with correlated observations. Some comments on the NDBM's are given in 4.5 and finally a short summary of the chapter is given in 4.6.

## 4.2. NORMAL WEIGHTED BAYESIAN MODELS :

Consider a parameterised process  $\{Y_t | \theta_t\}$ , where  $Y_t$  is an observable vector and  $\theta_t$  is the unobservable vector of state parameters containing certain process state characteristics of interest. For example each component of  $Y_t$  may represent the sales of

a product at time  $t$  with the corresponding component of  $\theta_t$ , representing its level of demand at that time. Each of the  $Y_t$  and  $\theta_t$  are random vectors and so, have probability distributions. Although the discount principle discussed in Chapter 2 provides an operationally simple and efficient method of estimation, no distributional assumption is made for  $\theta_t$ . This drawback can be overcome simply by introducing an initial joint probability distribution for  $Y_0$  and  $\theta_0$  and using Bayes theorem for updating the distribution of the parameters as new data arrives. This requires a model which describes the way that the parameters evolve with time and the amount of precision lost at each transition stage. The relevant model assumptions are stated in (3.1) and (3.2) for the DLM.

In order to introduce the discount principle into the Bayesian approach, the class of Normal Weighted Bayesian Models ( NWBM ) is defined

#### DEFINITION

For a parameterised process  $\{Y_t|\theta_t\}$ ,  $t > 0$ , a NWBM is given by a quadruple  $\{F, G, V, H\}_t$ , where

the observation probability distribution is

$$(Y_t|\theta_t) \sim N[F_t\theta_t; V_t] \quad (4.1)$$

and given the posterior parameter distribution at time  $t-1$

$$(\theta_{t-1}|D_{t-1}) \sim N[m_{t-1}; C_{t-1}] , \quad (4.2)$$

the prior parameter distribution at time  $t$ , is

$$(\theta_t|D_{t-1}) \sim N[G_t m_{t-1}; R_t] ; R_t = H_t C_{t-1} H'_t \quad (4.3)$$

Note that,  $R_t$  is a variance matrix provided that  $C_{t-1}$  is.  $C_{t-1}$  and  $V_t$  are variance matrices by definition.

#### THEOREM 4.1.

For any NWBM, the one step ahead forecasting distribution and the updating



parameter distributions are given by:

$$(Y_t | D_{t-1}) \sim N[\hat{y}_t; \hat{Y}_t] , \quad (4.4)$$

$$(\theta_t | D_t) \sim N[m_t; C_t] \quad (4.5)$$

where  $D_t = \{y_t, F_t, G_t, D_{t-1}\}$  and

$$\hat{y}_t = F_t G_t m_{t-1} , \quad \hat{Y}_t = F_t R_t F_t' + V_t , \quad (4.6)$$

$$m_t = G_t m_{t-1} - A_t e_t , \quad e_t = y_t - \hat{y}_t \quad (4.7)$$

$$C_t = (I - A_t F_t) R_t , \quad A_t = R_t F_t' \hat{Y}_t^{-1} = C_t F_t' V_t^{-1} . \quad (4.8)$$

## PROOF

The proof is standard in normal Bayesian theory. However, the results can be obtained from the identity

$$f(Y_t | \theta_t) f(\theta_t | D_{t-1}) = f(Y_t | D_{t-1}) f(\theta_t | Y_t, D_{t-1})$$

where the functions  $f(\cdot)$ 's are density functions of the appropriate random variables. Similarly, by rearranging the quadratic terms

$$(Y_t - F_t \theta_t)' V_t^{-1} (Y_t - F_t \theta_t) + (\theta_t - G_t m_{t-1})' R_t^{-1} (\theta_t - G_t m_{t-1})$$

as

$$(Y_t - \hat{y}_t)' \hat{Y}_t^{-1} (Y_t - \hat{y}_t) + (\theta_t - m_t)' C_t^{-1} (\theta_t - m_t) ,$$

where  $\hat{y}_t$ ,  $\hat{Y}_t$ ,  $m_t$  and  $C_t$  are as defined in the theorem.

NWBM's form an extensive class of models containing linear and non-linear models for which the prior and posterior distributions are normal. If  $W_t = H_t C_{t-1} H_t' - G_t C_{t-1} G_t'$  is nonnegative definite, the conditional distribution  $(\theta_t | \theta_{t-1}) \sim N[G_t \theta_{t-1}; W_t]$  may be introduced to provide coherent lead time forecast distributions. Thus, under this condition, any NWBM is a normal DLM. On the other hand, setting  $H_t = (G_t C_{t-1} G_t' + W_t)^{-1} C_{t-1}^{-1}$ , it is evident that any normal DLM

these different model components.

Before introducing practically more efficient and parsimonious NDBM settings, it is interesting to point out some relations with other well known models.

#### THEOREM 4.2.

Given a NDBM  $\{F, G, V, B\}_t$ , with non singular  $C_t$  for all  $t$ , and initial setting  $(m_0; C_0)$ ,

- i- If  $B_t = \beta I$ , and  $V_t = V$ , the NDBM forecast function is identical to that of of EWR  $\{F_t, G_t, \beta\}$  with initial setting  $(m_0; Q_0 = C_0^{-1} V)$
- ii- The NDBM forecast function is identical to that of DWE  $\{F, G, B\}_t$  with initial setting as in (i) with  $V_t = V$ .
- iii- If  $B_t = I$ , the joint posterior parameter distributions are identical to those of the DLM  $\{F, G, V, 0\}_t$ , with the same initial settings.

#### PROOF

The proof is by induction. From the assumptions, since in all the cases, common  $F_t$  and  $G_t$  are taken, the theorem is true for  $t=0$ . Now, assuming that the theorem is true at time  $t-1$ , we show that it is true for time  $t$ ,

- i- Given  $[m_{t-1}; C_{t-1}]$ , from the NDBM results, for time  $t$ , we have

$$m_t = G_t m_{t-1} + a_t e_t,$$

$$C_t^{-1} = R_t^{-1} + F'_t V^{-1} F_t; \quad R_t^{-1} = \beta G'_t{}^{-1} C_{t-1}^{-1} G_t^{-1}$$

This gives

$$C_t^{-1} V = \beta G'_t{}^{-1} Q_{t-1} G_t^{-1} + F'_t F_t = Q_t \quad \text{for EWR.}$$

Hence,

$$a_t = C_t^{-1} F'_t V = Q_t F'_t \quad \text{as for the EWR.}$$



these different model components.

Before introducing practically more efficient and parsimonious NDBM settings, it is interesting to point out some relations with other well known models.

#### THEOREM 4.2.

Given a NDBM  $\{F, G, V, B\}_t$ , with non singular  $C_t$  for all  $t$ , and initial setting  $(m_0; C_0)$ ,

- i- If  $B_t = \beta I$ , and  $V_t = V$ , the NDBM forecast function is identical to that of of EWR  $\{F_t, G_t, \beta\}$  with initial setting  $(m_0; Q_0 = C_0^{-1} V)$
- ii- The NDBM forecast function is identical to that of DWE  $\{F, G, B\}_t$  with initial setting as in (i) with  $V_t = V$ .
- iii- If  $B_t = I$ , the joint posterior parameter distributions are identical to those of the DLM  $\{F, G, V, 0\}_t$ , with the same initial settings.

#### PROOF

The proof is by induction. From the assumptions, since in all the cases, common  $F_t$  and  $G_t$  are taken, the theorem is true for  $t=0$ . Now, assuming that the theorem is true at time  $t-1$ , we show that it is true for time  $t$ ,

- i- Given  $[m_{t-1}; C_{t-1}]$ , from the NDBM results, for time  $t$ , we have

$$m_t = G_t m_{t-1} + a_t e_t,$$

$$C_t^{-1} = R_t^{-1} + F'_t V^{-1} F_t; \quad R_t^{-1} = \beta G'_t{}^{-1} C_{t-1}^{-1} G_t^{-1}$$

This gives

$$C_t^{-1} V = \beta G'_t{}^{-1} Q_{t-1} G_t^{-1} + F'_t F_t = Q_t \quad \text{for EWR.}$$

Hence,

$$a_t = C_t^{-1} F'_t V = Q_t F'_t \quad \text{as for the EWR.}$$

ii- The proof is similar to part one, with  $R_t^{-1} = B_t^{-1} G_t^{-1} C_{t-1}^{-1} G_t^{-1} B_t$ .

iii- The two models are identical, since, for the NDBM with  $B=I$ ,

$$R_t = G_t C_{t-1} G_t' = G_t C_{t-1} G_t' + 0 \text{ for the DLM with } W_t = 0.$$

Furthermore it can be seen from the Theorem in Section 3.3, that the limiting forecast distribution of a constant NDBM  $\{F, G, V, B\}$ , is identical to that of a constant DLM  $\{F, G, V, W\}$  with  $W = (HCH' - CGG')V$  being nonnegative definite. However, as with the DWE specifications for the time series models defined in Section 2.3.2, particular emphasis is given to canonically structured NDBM's for which  $G_t = \text{diag}\{G_1, G_2, \dots, G_r\}_t$  with  $G_i$  of full rank  $n_i$ .  $B_t = \text{diag}\{\beta_1 I_1, \beta_2 I_2, \dots, \beta_r I_r\}$  where  $I_i$  is the identity matrix of dimension  $n_i$  and  $0 < \beta_i \leq 1$  :  $i=1, 2, \dots, r$ .

#### 4.3.2. Forecasting and Updating with NDBM's

Given the posterior parameter distribution at time  $t-1$  as in (4.2), the one step ahead forecast distribution and the updating posterior parameter distributions at time  $t$ , are given by (4.4) and (4.5), with  $R_t = B_t^{-1} G_t C_{t-1} G_t' B_t^{-1}$ .

The  $k$ -steps ahead forecast function  $F_t(k)$ ,  $k > 0$ , is given by

$$F_t(k) = E[Y_{t+k} | D_t] = F_{t+k} \left[ \prod_{i=1}^k G_{t+i} \right] m_t$$

#### 4.3.3. Coherency

For the NWBM and NDBM's coherent joint forecast distributions may be derived using a corresponding DLM

$$\begin{aligned} Y_{t+k} &= F_{t+k} \theta_{t+k} + v_{t+k}, & v_{t+k} &\sim N[0; V_{t+k}] \\ \theta_{t+k} &= G_{t+k} \theta_{t+k-1} + w_{t,k}; & w_{t,k} &\sim N[0; W_{t,k}]. \end{aligned}$$

Defining ,

$$R_{t,1} = H_{t+1} C_t H_{t+1}'$$

$W_{t,k} = \{w_{i,j}\}$  and  $R_{t,k} = \{r_{i,j}\}$ ,  $W_{t,k}$  is derived from the recursive relationships.

$$w_{i,j} = (1 - (\beta_i \beta_j)^2) r_{i,j} \quad (4.9)$$

$$R_{t,k+1} = H_{t+k+1} (I - A_{t,k} F_{t+k}) R_{t,k} H'_{t+k+1}$$

$$A_{t,k} = R_{t,k} F'_{t+k} (V_{t+k} - F_{t+k} R_{t,k} F'_{t+k})^{-1}$$

$$H_{t-k} = B^{-1} G_{t-k}$$

For univariate series,  $V_{t-k} + F_{t-k} R_{t,k} F'_{t+k}$  is a scalar quantity. Hence, the calculation of  $W_{t,k}$  does not require matrix inversions.

### THEOREM 4.3.

Given no missing observations, the NDBM  $\{F, G, V, B\}$  is coherent.

### PROOF

Let  $(\theta_t | D_t) \sim N[m_t; C_t]$ .

Define

$$R = G C_t G' + W_{t+1} = B^{-1} G C_t G' B'^{-1}$$

$$Q = G R G' + W_{t+2} = B^{-1} G R G' B'^{-1}$$

$$\hat{Y} = F R F' + V, \quad \hat{Z} = F R F' + V.$$

Note that the calculation of  $W_{t+2}$  is on the basis that  $y_{t+1}$  is missing. In the above relations, some subscripts are removed for convenience.

Using the formal DLM relations,

$$\begin{pmatrix} Y_{t+2} \\ Y_{t+1} \\ \theta_t \\ \theta_{t+1} \\ \theta_{t+2} \end{pmatrix} | D_t \sim N \left[ \begin{pmatrix} F G^2 m_t \\ F G m_t \\ m_t \\ G m_t \\ G^2 m_t \end{pmatrix}; \begin{pmatrix} \hat{Z} & F G R F' & F G^2 C_t & F G R & F Q \\ \cdot & \hat{Y} & F G C_t & F R & F R G' \\ \cdot & \cdot & C_t & G C_t & G^2 C_t \\ \cdot & \cdot & \cdot & R & G R \\ \cdot & \cdot & \cdot & \cdot & Q \end{pmatrix} \right].$$

It can be seen that the posterior distribution of  $(\theta_{t+1}|D_{t+1})$  is

$$N[\mathbf{m}_{t+1} = \mathbf{G}\mathbf{m}_t + \mathbf{R}\mathbf{F}'\hat{\mathbf{Y}}^{-1}(\mathbf{y}_{t+1} - \mathbf{F}\mathbf{G}\mathbf{m}_t); \mathbf{C}_{t+1} = \mathbf{R} - \mathbf{R}\mathbf{F}'\hat{\mathbf{Y}}^{-1}\mathbf{F}\mathbf{R}]$$

But

$$(\theta_{t+2}|\mathbf{y}_{t+1}, D_t) \sim N[\mathbf{G}\mathbf{m}_{t+1}; \mathbf{Q} - \mathbf{G}\mathbf{R}\mathbf{F}'\hat{\mathbf{Y}}^{-1}\mathbf{F}\mathbf{R}\mathbf{G}]$$

Now  $\mathbf{Q} - \mathbf{G}\mathbf{R}\mathbf{F}'\hat{\mathbf{Y}}^{-1}\mathbf{F}\mathbf{R}\mathbf{G} \neq \mathbf{B}^{-1}\mathbf{G}\mathbf{C}_{t-1}\mathbf{G}'\mathbf{B}^{-1}$  showing that the discount principle would be incoherent. However, for the  $W_{t,k}$ 's defined by (4.9), since  $\mathbf{C}_{t,k} = \mathbf{C}_{t-k}$ , this inequality will not occur and the above procedure can be extended to establish the equivalence of the DLM and the NDBM for  $\mathbf{Y}_{t-k}$ .

The above theorem ensures the testability of the NWBM's on the same lines with the DLM. However, given a starting initial prior  $[\mathbf{m}_0; \mathbf{C}_0]$  together with  $\mathbf{F}, \mathbf{G}, \mathbf{V}$  and  $\mathbf{B}$ , successive predictive distributions can be used to generate data sets following particular NWBM's. The noted difference is that the DLM uses the set of equations (3.1) and (3.2) assuming that  $\theta_0$  is known while, the NWBM starts with  $[\mathbf{m}_0; \mathbf{C}_0]$  and the transition uncertainty is acknowledged by the discount matrix  $\mathbf{B}$ .

#### 4.3.4. Sequential analysis of designed experiments :

Statistical experiments are often performed sequentially and are subject to slow movements as well as sharp changes, perhaps due to some uncontrollable sources of variation. In such cases, static ( non sequential ) models are hardly justified and may well lead to false conclusions. DLM's have been adopted for a sequential analysis of quality characteristic in the production of Nylon Polymer by Harrison. However, the problems already discussed regarding the  $\mathbf{W}$  covariance matrix often arise. The NDBM's overcome these problems. For example, a  $2^2$  completely randomised sequential experiment can be represented by

$$Y_{ijt} = \theta_{1,t} + \theta_{ijt} + \epsilon_{ijt} \quad i, j = 1, 2$$

where  $\theta_{1,t}$  represents the block effect and  $\theta_{ijt}$  represents the collective treatments effect with  $\sum_i \sum_j \theta_{ijt} = 0$  at any time stage  $t$ .



Now, in order to partition the variation among the treatments, an orthogonal contrast need to be constructed by partitioning  $\theta_{ijt}$  to  $\theta_{2,t}$ ,  $\theta_{3,t}$  and  $\theta_{4,t}$  where  $\theta_{1,t}$  represents the block effect,  $\theta_{2,t}$  and  $\theta_{3,t}$  as the effects of treatment 1 and 2 respectively and  $\theta_{4,t}$  is the interaction effect. Usually this is performed as follows:

- i- the effect in presence of both treatments and their interaction  $\propto$  sum of the treatments 1 and 2 + random error.

$$Y_{11t} = \theta_{1,t} - \theta_{2,t} - \theta_{3,t} - \theta_{4,t} - \epsilon_{1t}$$

- ii- main effect of treatment 1  $\propto$  sum of the terms with treatment 1 - sum of the terms without treatment 1 + random error.

$$Y_{12t} = \theta_{1,t} + \theta_{2,t} - \theta_{3,t} - \theta_{4,t} + \epsilon_{2t}$$

- iii- similarly, for the main effect of treatment 2, we have

$$Y_{21t} = \theta_{1,t} - \theta_{2,t} + \theta_{3,t} - \theta_{4,t} + \epsilon_{3t}$$

The orthogonality condition suggests that

$$Y_{22t} = \theta_{1,t} - \theta_{2,t} - \theta_{3,t} + \theta_{4,t} + \epsilon_{4t}$$

In collecting the above information with  $Y'_t = [Y_{11}, Y_{12}, Y_{21}, Y_{22}]_t$ , an appropriate NDBM may be  $\{F, I, V, B\}$  where

$$F = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad (4.10)$$

$G$  is taken as the identity matrix to indicate a steady parameter evolution with time.

#### 4.4. OTHER IMPORTANT SPECIAL CASES:

##### 4.4.1. The Modified NDBM

In modelling discontinuities and changes in different characteristics of processes using intervention or multiprocess models it is often advisable to operate a system that



protects the information on unchanged components against unwanted interaction effects arising from those components that have been disrupted. This is possible and the occurrence of discontinuities in the data need not require complete respecification of the parameters as has often been suggested by statisticians such as Jaynes (1983).

## DEFINITION

Let  $(\mathbf{0}_{t-1} | D_{t-1}) \sim N[\mathbf{m}_{t-1}; \mathbf{C}_{t-1}]$ ,  $\mathbf{G} = \text{diag} \{ \mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_r \}$ ,  $\mathbf{B} = \text{diag} \{ \mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_r \}$ , and let the partitioned structures of  $\mathbf{C}_{t-1}$  be  $\{\mathbf{C}_{i,j}\}_{t-1}$ , and of  $\mathbf{R}_t$  be  $\{\mathbf{R}_{i,j}\}_t$ , for  $i, j = 1, 2, \dots, r$ . A modified NDBM is a NDBM  $\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{B}\}_t$  such that

$$\mathbf{R}_{ii} = \mathbf{B}_i^{-1} \mathbf{G}_i \mathbf{C}_{ii} \mathbf{G}_i' \mathbf{B}_i^{-1}, \quad i = 1, 2, \dots, r. \quad (4.11)$$

$$\mathbf{R}_{i,j} = \mathbf{G}_i \mathbf{C}_{i,j} \mathbf{G}_j', \quad \text{for } i \neq j \quad (4.12)$$

The occurrence of sudden structural changes in the state of sequential statistical processes is common. In time series processes, it may be possible to classify the types of change into changes in level, growth and/or seasonal components. Such changes can be modelled by increasing uncertainty only to the corresponding components so that, other components uncertainty is not effected. In DLM's, this is performed by increasing uncertainty of the state error vector,  $\mathbf{w}_t$ , only for the relevant blocks. For the NDBM's, future uncertainty is controlled by the discount matrix. It can be observed from the definition of NDBM, that the future uncertainty introduced to a particular block will be transmitted to other blocks through their correlation between them. The modified NDBM is introduced to prevent this. Moreover, a major disturbance on a particular block may be signaled with intervention using a discount factor  $\beta^N$ , where  $N$  is chosen to age the effect of past history relevant to that component by  $N$  periods. This can be performed even within blocks. Although, this loses invariance under linear transformations temporarily, but enables to introduce uncertainty into a desired component of that block. These ideas are used in the examples that are given in Chapter 8 and Migon and Harrison (1983) have applied modified NDBM's in their models.

The above definition can be modified to include more general transition  $G$  matrices. That is, given the partitioned form of  $G$  as  $\{G_{ij}\}$  and that of  $GC_{t-1}G' = \Sigma$  as  $\{\Sigma_{ij}\}$ , (4.11) and (4.12) are replaced by

$$R_{ii} = B_i^{-1} \Sigma_{ii} B_i^{-1}, i=1,2,\dots,r \text{ and}$$

$$R_{ij} = \Sigma_{ij} \text{ for } i \neq j.$$

#### 4.4.2. Extended NDBM's

In many applications an NDBM provides an adequate model but other applications may require a more general NWBM. This is particularly the case when  $G$  is singular and when high frequencies and some type of stochastic transfer responses are to be modelled. The extended NDBM is defined by the quintuple  $\{F, G, V, B, W\}_t$  where, given  $(\theta_{t-1} | D_{t-1}) \sim N[m_{t-1}; C_{t-1}]$ , this defines

$$(Y_t | \theta_t) \sim N[F_t \theta_t; V_t]$$

$$(\theta_t | D_{t-1}) \sim N[G_t m_{t-1}; R_t],$$

where

$$R_t = B_t^{-1} G_t C_{t-1} G_t' B_t^{-1} + W_t$$

Often in regression and design of experiments, some of the variables are fairly stable with a constant variance and it may not be advisable to subject their precision to an exponential decay. This may be the case with the example in 4.3.4, if the block effects are independent or exchangeable so that  $(\theta_{1,t} | D_{t-1}) \sim N[\theta_{5,t}; \sigma^2]$  where  $\theta_{5,t}$  is unknown and either static or subject to a very slow movement. With the design matrix  $F$  defined by (4.10), this may be modelled using an extended NDBM with

$$F_1 = [F \ 0], \quad G = \begin{bmatrix} 0 & (0,0,0,1) \\ 0 & I_4 \end{bmatrix},$$

$\beta_1 = \beta_5$  and  $W = \{W_{i,j}\}$  with  $W_{1,1} = \sigma^2$  its only non zero element.

Similarly in modelling correlated observation errors such as those generated by a stationary second order autoregressive process  $v_t = (1-\phi_1 B)(1-\phi_2 B)\delta_t$ , models of type

$$\{[1, 0, \mathbf{0}_n]; \begin{bmatrix} \phi_1 & 1 & \mathbf{0} \\ 0 & \phi_2 & \mathbf{F}_{t+2} \\ \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix}; 0; \text{diag}\{\{1, 1, \mathbf{B}\}, \mathbf{W}\} \quad \text{m}$$

may be preferred. The only non zero element of  $\mathbf{W}$  is  $W_{22} = V$  and this can be easily estimated on line. See also Zellner (1971) for autoregressive models. For a particular extension of EWR to Generalised EWR which considers stationary observation errors see Harrison and Akram (1983).

#### 4.5. SUMMARY:

The discount concept is introduced into Bayesian modelling and forecasting via a general class of NWBM models. Some special important and parsimonious subclasses of NDBM's , Modified NDBM's and the Extended NDBM's are introduced. Particular attention is given to the NDBM's. Neat updating formulas for the location vector and scale matrices are derived together with their forecast distributions.



## CHAPTER FIVE

### ON-LINE VARIANCE LEARNING

#### 5.1. INTRODUCTION:

One of the consequences of the conceptual differences between the Bayesian representation of a time series as a Markovian process and its non Bayesian formulation is that the former allows for a genuine dynamic structure for the variance  $V_t$  of the observation error  $v_t$  that is often assumed to be a known constant in the formulations of Kalman Filtering and ARIMA techniques.

On-line estimation of  $V_t$  is important for a successful practical application of Bayesian forecasting but it is crucial in multiprocess modelling since it governs the likelihoods of the different models. Experience has shown that practitioners have little intuitive feeling for the size of this variance. It is often confused with the one step ahead forecast variance  $\hat{Y}_t$ . However for single constant NDBM cases, in the period of stability, the relationship is  $V = \prod_{i=1}^n (\beta_i / \lambda_i^2) \lim_{t \rightarrow \infty} \hat{Y}_t$  ( part ii of Theorem 6.4.) , where  $\beta_i$  is the discount factor associated with the  $i^{th}$  parameter  $\theta_i$  with associated eigenvalue  $\lambda_i$ . If required the extra uncertainty associated with  $V$  may be acknowledged and used to derive marginal forecast distributions . This is easily done since  $(\theta, V)$  is jointly estimated in a neat Bayesian manner.

A number of approaches based on the idea of De Groot(1970) have been adopted for estimating the observation variance  $V_t$ . Smith (1977) has discussed the problem for univariate steady state models. The case of heavy tailed error distributions is given by West (1982). The Bayesian procedure introduced in 5.2 can also be seen in Ameen and Harrison (1983 c). Other non Bayesian elaborations introduced by Harrison and Stevens(1975), Harrison and Pearce(1972) and Cantarelis and Johnston(1983) are briefly



reviewed and a generalisation of the latter is given in 5.3. A new procedure called the power law is given in 5.4, Ameen and Harrison (1983 b).

## 5.2. THE BAYESIAN APPROACH:

It is assumed that the variance  $V_t = I\phi_t^{-1}$  where  $\phi_t$  is unknown.

The observation distribution is

$$(Y_t | \theta_t, \phi_t) \sim N[F_t \theta_t; I\phi_t^{-1}], \quad (5.1)$$

The posterior state distributions are :

$$(\theta_{t-1} | D_{t-1}, \phi_{t-1}) \sim N[m_{t-1}; C_{t-1}\phi_{t-1}^{-1}] \quad (5.2)$$

$$(\phi_{t-1} | D_{t-1}) \sim \Gamma\left(\frac{\alpha_{t-1}}{2}, \frac{\eta_{t-1}}{2}\right) \quad (5.3)$$

with this Gamma pdf having a kernel

$$\exp\{((\eta_{t-1}/2) - 1)\log\phi_{t-1} - (\alpha_{t-1}/2)\phi_{t-1}\} \quad (5.4)$$

Defining the prior pdf's as

$$(\theta_t | D_{t-1}, \phi_t, I_t) \sim N[Gm_{t-1}; R_t\phi_t^{-1}] \quad (5.5)$$

$$(\phi_t | D_{t-1}, I_t) \sim \Gamma(\mu(\alpha_{t-1})/2; \lambda(\eta_{t-1})/2) \quad (5.6)$$

where  $I_t$  represents the information required in the posterior to prior transition,  $R_t = H_t C_{t-1} H_t'$  and  $\mu(\cdot), \lambda(\cdot)$  are feasible functions (such that (5.6) is well defined) of the posterior parameters respectively.

The functions  $\mu$  and  $\lambda$  can play an important role in both theory and applications. A special choice is introduced later in Section 5.4. Other forms are dealt with in Ameen (1983 b). These are specific functions either defined through posterior entropies or accommodate some external information like advertising awareness. However, it follows from (5.1)-(5.6) and using Bayes theorem, that the recurrence relationships for  $m, C, \hat{y}_t$  and  $\hat{Y}_t$  are exactly as (4.6)-(4.8) with the setting  $V=I$ , and

$$(Y_t | D_{t-1}, I_t, \phi_t) \sim N[\hat{y}_t; \hat{Y}_t \phi_t^{-1}]$$

$$(\theta_t | D_{t-1}, \phi_t) \sim N[\bar{m}_t; C_t \phi_t^{-1}]$$

$$(\phi_t | D_t) \sim \Gamma(\alpha_t/2; \eta_t/2)$$

where  $\eta_t = \lambda(\eta_{t-1}) + 1$  and  $\alpha_t = \mu(\alpha_{t-1}) + e_t' \hat{Y}_t^{-1} e_t$ . As usual  $e_t = y_t - \hat{y}_t$ . The joint distribution  $(Y_{t+1}, \phi_{t+1} | D_t)$  is readily obtained and  $(Y_{t-1} | D_t)$  is derived by integrating out  $\phi_{t+1}$ . In the univariate case

$$\frac{(Y_{t+1} - \hat{y}_{t+1})}{(\hat{Y}_{t+1} \phi_{t+1})^{1/2}} \sim t_{\eta_t},$$

the student  $t$ -distribution with  $\eta_t$  degrees of freedom.

This method is operationally elegant and is properly Bayesian. It is not easy to retain the elegance when generalising to many cases where the correlation structure of  $V_t$  is unknown or where  $V_t$  is not a constant. Consequently practitioners may prefer the robust variance estimation method discussed in 5.4

### 5.3. NON BAYESIAN METHODS - A SHORT REVIEW:

In addition to the method mentioned in 5.2, a number of approaches have been adopted for estimating the observation variance  $V$ . Harrison and Pearce (1972) used a six point curve fitting around each data point. Denoting the value of the curve at that point by  $\hat{y}_t$  and assuming that  $y_t - \hat{y}_t \sim N[0; V_t]$  where  $V_t \propto \hat{y}_t^{2\beta}$ , they chose the maximum likelihood estimate of  $\beta$ . Another method proposed by Harrison and Stevens (1975) assumes that  $\sqrt{V_t} \propto L_t^P S_t^Q$  where  $L_t, S_t$  are the level and seasonality components and  $P, Q$  are known constants while the proportionality constant  $C$  (say) is obtained from the median of a pre specified  $N$  ordered constants with corresponding probabilities which are updated on line using the data information. These methods are not theoretically profound and do not generalise easily. Another on-line estimation method based on the limiting properties of the steady state DLM's is suggested by Cantarelis and Johnston (1983). This is described as follows

$$\hat{V}_t = \left( \frac{t-1}{t} \right) \hat{V}_{t-1} + \frac{1-a}{t} e_t^2$$

where  $a = \lim_{t \rightarrow \infty} a_t$ ,  $a_t$  being the adaptive coefficient.

A direct generalisation of this method can be given as

$$\hat{V}_t = \frac{t-1}{t} \hat{V}_{t-1} + \frac{1}{t} (1 - f_t a_t) e_t^2$$

or more generally

$$\hat{V}_t = (1-\beta) \hat{V}_{t-1} + \beta (1 - f_t a_t) e_t^2$$

#### 5.4. THE POWER LAW:

A more general but simple, efficient and robust procedure can be described using the relationship  $V_t = (I - P_t A_t) \hat{Y}_t$ .

For a univariate time series, define  $d_t^2 = (1 - f_t a_t) e_t^2$ . In parallel with the Bayesian approach described in Section 5.2, the estimate of  $V_t$  may be given by

$$\hat{V}_t = X_t / \eta_t$$

where

$$X_t = X_{t-1} + d_t^2 \quad (5.7)$$

$$\eta_t = \eta_{t-1} + 1 \quad (5.8)$$

Initially  $(X_0, \eta_0)$  may be chosen such that  $\hat{V}_0 = X_0 / \eta_0$  is a point estimate for  $V_0$  and  $\eta_0$  is the accuracy expressed in terms of degrees of freedom, or of equivalent observations. In the analysis of 5.2 it is seen that  $\hat{V}_t = 1/E[\phi|D_t]$ . Hence, if required, forecasts can be produced as in 5.2 using a student t-distribution with  $\eta_t$  degrees of freedom. In practice it may be wise to protect the estimate from outliers and major disturbances. Outlier-prone distributions, O'Hagan (1979) can be introduced using mixture distributions. However one simple effective practical method is to define  $d_t^2 = (1 - f_t a_t) \min\{e_t^2, K \hat{Y}_t\}$  where in general the constant  $K$  belongs to the interval  $[4, 6]$ . In those cases in which it is



suspected that  $V_t$  varies slowly over time a discount factor  $\beta$  may be introduced by replacing (5.7) and (5.8) by

$$X_t = \beta X_{t-1} + d_t^2, \quad \text{and}$$

$$\eta_t = \beta \eta_{t-1} + 1$$

This procedure is easily applied and experience with both pure time series and regression type models is encouraging. However, because of the skew distribution of  $d_t^2$  it is wise to choose  $0.95 < \beta < 1$ . Further if the initial prior of the parameter vector  $\theta$  is vague, then it is recommended that variance learning commences at time  $n+1$  where  $n$  is the dimension of the state vector. In stock control, with positive observations, an empirical variance law  $V_t = a \hat{y}_t^{2b}$  with  $b = 0.75$  is often used. Stevens(1974). An estimate  $\hat{a}_t$  of  $a$  is then derived as

$$\hat{a}_t = Z_t / \eta_t, \quad \text{where}$$

$$Z_t = \beta Z_{t-1} + d_t^2 / \hat{y}_t^{1.5}$$

$$\eta_t = \beta \eta_{t-1} + 1$$

Future estimates of  $V$  are;  $\hat{V}_{t+k} = \hat{a}_t \{E(Y_{t+k} | D_t)\}^{1.5}$ .

A more general procedure for accommodating stochastic scale parameters is as follows. See Ameen (1983 c).

Let  $\phi_t$  be a scale parameter with posterior probability density function (pdf) at time  $t-1$  given by (5.3) and prior pdf for time  $t$ , be given by (5.6). Moreover, let  $x_t$ ,  $m_{t-1}$  and  $k_t$  be the modes of the random variables with pdf's  $f(Y_t | \theta_t)$ ,  $f(\theta_{t-1} | D_{t-1})$  and  $f(\theta_t | D_{t-1})$  respectively.

Define the link between  $Y_t$ ,  $\theta_t$  and  $\phi_t$  as follows

$$f(Y_t | \theta_t, \phi_t) \propto \phi_t^{1-\phi_t} f^{1-\phi_t}(Y_t | x_t) f^{\phi_t}(Y_t | \theta_t) \quad (5.9)$$

$$f(\theta_t | \phi_t, D_{t-1}) \propto \phi_t^{1-\phi_t} f^{1-\phi_t}(k_t | D_{t-1}) f^{\phi_t}(\theta_t | D_{t-1}) \quad (5.10)$$

Combining ( 5.6 ) with ( 5.9 ) and ( 5.10 ), the approximate kernel of the posterior pdf for  $(\theta_t, \phi_t | \mathbf{y}_t, D_{t-1})$  is

$$\begin{aligned} & \phi_t^{\lambda(\eta_{t-1})/2} e^{-\phi_t \mu(\alpha_{t-1})/2} [f(\mathbf{y}_t | \mathbf{x}_t) f(\mathbf{h}_t | D_{t-1})]^{1-\phi_t} [f(\mathbf{y}_t | \theta_t) f(\theta_t | D_{t-1})]^\phi \\ & \propto \phi_t^{\lambda(\eta_{t-1})/2} e^{-\phi_t \mu(\alpha_{t-1})/2} [f(\mathbf{y}_t | \mathbf{x}_t) f(\mathbf{h}_t | D_{t-1}) / f(\mathbf{m}_t | D_t)]^\phi f^{1-\phi}(\mathbf{m}_t | D_t) f^\phi(\theta_t | D_t) \\ & = \phi_t^{\lambda(\eta_{t-1})/2} \exp\{-[\mu(\alpha_{t-1}) + 2 \ln \frac{f(\mathbf{y}_t | \mathbf{x}_t) f(\mathbf{h}_t | D_{t-1})}{f(\mathbf{m}_t | D_t)}] \phi_t / 2\} f^{1-\phi}(\mathbf{m}_t | D_t) f^\phi(\theta_t | D_t) \end{aligned}$$

where  $\mathbf{m}_t$  is the posterior mode of  $\theta_t | D_t$ . In comparison with the approximate posterior pdf

$$f(\theta_t, \phi_t | D_t) \propto \phi_t^{\lambda(\eta_{t-1})/2} e^{-\phi_t \mu(\alpha_{t-1})/2} f^{1-\phi}(\mathbf{m}_t | D_t) f^\phi(\theta_t | D_t) .$$

we have

$$\eta_t = \lambda(\eta_{t-1}) + 1$$

$$\alpha_t = \mu(\alpha_{t-1}) + 2 \ln \{ f(\mathbf{y}_t | \mathbf{x}_t) f(\mathbf{h}_t | D_{t-1}) / f(\mathbf{m}_t | D_t) \}$$

The formulas (5.9) and (5.10) are exact for normal random vectors and are in contrast with those of (5.1) and (5.5). However, the formulation above goes well beyond the exponential family of distributions and has the key for introducing a constructive dynamic evolution of location and scale parameters in generalised dynamic models.



## 5.5. SUMMARY:

A proper Bayesian on line estimation procedure for the observation variance is described in 5.2 . A short review of the existing non Bayesian techniques is given in 5.3. The power law is described in 5.4 . Outlines for a general model , for which stochastic scale parameters can be accommodated, is given.

## CHAPTER SIX

### LIMITING RESULTS

#### 6.1. INTRODUCTION :

There has been a continued interest in deriving limiting values for the parameter variance  $\mathbf{C}_t$  and the adaptive vector  $\mathbf{a}_t$  associated with observable constant DLM's but the difficulty in solving Riccati equations has restricted progress. However for constant NDBM's  $\{\mathbf{f}, \mathbf{G}, \mathbf{V}, \mathbf{B}\}$  these values can be obtained directly. Hence the results also apply to the set of constant DLM's  $\{\mathbf{f}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}$  which have limiting forecast distributions equivalent to those of constant NDBM's. These results are relevant to practice since convergence is often fast and, in order to achieve conceptual simplicity and parametric parsimony, previous efforts have been devoted to determining constant DLM's which have limiting forecast functions equivalent to those obtained by the application of EWR. Harrison and Akram (1983) and Roberts and Harrison (1984).

Similar models and the method of transforming from one similar model to another are defined. Limiting results for the state covariance matrix  $\mathbf{C}_t$  and the adaptive vector  $\mathbf{a}_t$  are stated first for models similar to a model with a diagonal transition matrix  $\mathbf{G}$  and then for general constant NDBM's. The limiting relationship between the observations and the one step ahead prediction errors is obtained for NDBM's. This leads to the establishment of a relationship between the ARIMA models and the constant NDBM's.

#### 6.2. SIMILAR MODELS AND REPARAMETRISATION :

One of the desirable objectives of theoretical developments is to obtain unified results that may be used in different fields of applications. By looking at the most meaningful economical representations, this eases the understanding of practitioners. In NDBM's this leads to canonical representations of categorised models. The properties of other more complicated models can be studied through their similarity with the canonical models. Harrison and Akram (1983) have discussed reparametrisations within the class of

DLM's.

## DEFINITION

A constant NWBM  $\{F, G, V, H\}$  is called observable if 
$$\begin{bmatrix} F \\ FG \\ FG^2 \\ \vdots \\ FG^{n-1} \end{bmatrix}$$
 is of full rank.

The observability condition for NWBM's is to ensure the estimability of state parameter vectors at any time stage from a finite number of observations.

## DEFINITION

Two NWBM's  $M_i = \{F_i, G_i, V, H_i\}$   $i=1,2$  are said to be similar if there exists a non singular transformation  $L$  such that

$$\{F_1 L^{-1}, L G_1 L^{-1}, V, L H_1 L^{-1}\} = \{F_2, G_2, V, H_2\}$$

The importance of finding similar models arise in practice since, real life problems are rather complicated in their 'primary' statistical formulation. Apart from computational benefits, this provides physically meaningful relationships among the primary variables and those of the canonical form like growth, level and seasonality components.

## THEOREM 6.1.

If  $M_1$  and  $M_2$  are two observable similar NWBM's then  $L = T_2^{-1} T_1$  where

$$T_i = [F_i', (F_i G_i)', \dots, (F_i G_i^{n-1})']', \quad i=1,2$$

## PROOF

Since  $M_1$  and  $M_2$  are similar, it follows that  $F_2 = F_1 L^{-1}$  and  $G_2^k = L G_1^k L^{-1}$ ,

$$k = 1, 2, \dots, \{n \text{ minus } 1\}.$$

This gives

$$\begin{bmatrix} F_2 \\ F_2 G_2 \\ \vdots \\ F_2 G_2^{n-1} \end{bmatrix} = \begin{bmatrix} F_1 \\ F_1 G_1 \\ \vdots \\ F_1 G_1^{n-1} \end{bmatrix} L^{-1}$$

i.e:

$$T_2 = T_1 L^{-1}$$

From observability it follows that  $T_1$  and  $T_2$  are invertible. This gives

$$L = T_2^{-1} T_1$$

The above result introduces similar reparametrisations. That is if  $\theta$  is the state vector for the first model then the reparametrisation  $\phi = L\theta$  produces the model  $M_2$ . As forecast functions are characterised by the specifications of  $F$  and  $G$  and in particular the eigenvalues of  $G$  plays an important role in that specification, canonical forms are specifically useful in demonstrating these ideas.

### THEOREM 6.2.

Let  $\lambda_1, \lambda_2, \dots, \lambda_n$  be the eigenvalues of  $G$ , and  $0 < \beta/\lambda_i < 1$ ;  $i=1, 2, \dots, n$ . If the constant NDBM  $\{f, G, V, \beta I\}$  is observable, then

$$\lim_{t \rightarrow \infty} \{C, R, \hat{Y}, \hat{e}\}_t = \{C, R, \hat{Y}, \hat{e}\}$$

uniquely exists, with  $C$ , and  $R$  being non singular.

### PROOF

From the NDBM results,

$$Q_t = \beta G'^{-1} Q_{t-1} G^{-1} + f' f, \quad Q_t = C_t^{-1} V$$

$$Q_t = \beta^t G'^{-t} Q_0 G^{-t} + \sum_{i=0}^{t-1} \beta^i G'^{-i} f' f G^{-i} \quad (6.1)$$

Hence, using the assumptions  $0 < \beta^i/\lambda_i < 1$ ,  $\lim_{t \rightarrow \infty} Q_t = Q$  exist.

To show that  $Q$  is positive definite, consider (6.1) as  $t \rightarrow \infty$ , the first term converges to zero, and

$$Q = \sum_{i=0}^{\infty} \beta^i G'^{-i} f' f G^{-i} = \sum_{i=0}^{\infty} \beta^i G'^{-i} T T' G^{-i}$$

where  $T = [f', \beta G'^{-1} f', \dots, \beta^{n-1} G'^{-(n-1)} f']$ .

From observability,  $T = G'^{-(n-1)} [G'^{n-1} f', \beta G'^{n-2} f', \dots, \beta^{n-1} f']$  is non singular.

To show that  $Q$  is unique, assume that there exists  $Z$  such that  $Z = \beta G'^{-1} Z G^{-1} + f' f$ .

Therefore,

$$Z - Q = \beta G'^{-1} (Z - Q) G^{-1} \quad (6.2)$$

Successive applications of (6.2) gives  $Z - Q = \beta^k G'^{-k} (Z - Q) G^{-k}$ , and as  $k \rightarrow \infty$ , we have  $Z - Q = 0$ .

Since,  $Q_t = C_t^{-1} V$ ,  $R_t = \beta^{-1} G C_t G'$ ,  $\dot{Y} = f R_t f' + V$  and  $e_t = R_t f' \dot{Y}_t^{-1} = C_t f' V^{-1}$ , the limiting forms for  $C_t$ ,  $R_t$ ,  $\dot{Y}_t$  and  $e_t$  all exist and unique. Moreover,

$$Q = \beta G'^{-1} Q G^{-1} + f' f = C^{-1} f' V^{-1}, \quad R = \beta^{-1} G C G', \quad (6.3)$$

$$\dot{Y} = f R f' + V, \quad e = R f' \dot{Y}^{-1} = C f' V^{-1} \quad (6.4)$$

### THEOREM 6.3.

Let,  $\sum_{i=1}^r n_i = n$ ,  $G = \text{diag}\{G_1, G_2, \dots, G_r\}$  and  $B = \text{diag}\{\beta_1 I_1, \beta_2 I_2, \dots, \beta_r I_r\}$  with  $0 < \beta_i < \min\{1, |\lambda_{i,1}|^2, |\lambda_{i,2}|^2, \dots, |\lambda_{i,n_i}|^2\}$  where  $\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,n_i}$  are the eigenvalues of  $G_i$  with dimension  $n_i$ .



If the constant NDBM  $\{f, G, V, B\}$  is observable, then

$$\lim_{t \rightarrow \infty} \{C, R, \dot{Y}, a\}_t = \{C, R, \dot{Y}, a\}$$

uniquely exists. Moreover,  $C$  and  $R$  are non singular.

## PROOF

The proof is similar to that of Theorem (6.2), knowing that the observability of the model gives the observability in each model component block.

In order to have some insight into the sensitivity of these models, consider a DLM

$$Y_t = \theta_t + v_t \quad ; \quad v_t \sim N[0; V]$$

$$\theta_t = \theta_{t-1} + w_t \quad \sim N[0; W]$$

Given the prior  $(\theta_0 | D_0) \sim N[m_0; C_0]$ , it can be seen that the posterior state variance  $C_t$  and the adaptive coefficient  $A_t$ , both converge to  $C$  and  $A$  respectively and  $C = AV = ((W^2 + 4WV)^{-1} - W)/2$ .

Now, consider a NDBM with the same prior settings and take the discount factor  $\beta$  as  $1 - ((1 + 4V/W)^{-1/2} - 1)W/(2V)$ . This guaranties that the NDBM and the DLM both have the same limiting distribution. However, given any common posterior variance  $C_{t-1}$  at time  $t-1$ , the adaptive coefficient  $A_t(W)$  for the DLM is

$$A_t(W) = 1/(1 + V/(C_{t-1} + W))$$

while the alternative form under the NDBM is

$$A_t(\beta) = 1/(1 + \beta V/C_{t-1})$$

Therefore, if  $C_t > C$  for all  $t$ , the DLM converges faster than the NDBM to the limit. However, if  $C_t < C$  for all  $t$ , then the NDBM converges to the limit faster than the DLM. This generalises to higher dimensions.

### 6.3. A COMMON CANONICAL REPRESENTATION :

One of the most common and yet simple canonical forms for observable models with system matrices  $G$  which have distinct eigenvalues ,  $\lambda_1, \lambda_2, \dots, \lambda_n$  that is  $G = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  and  $f = [1, 1, \dots, 1]$ . For such an observable NDBM the following theorem holds

#### THEOREM 6.4.

Let  $\{f, G, V, B\}$  be a constant NDBM, with  $f = [1, 1, \dots, 1]$ ,  $G = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  and  $B = \text{diag}\{\beta_1, \beta_2, \dots, \beta_n\}$  where  $0 < \beta_i < |\lambda_i|^2$ ,  $u_i = \beta_i / |\lambda_i|$ ,  $i=1, 2, \dots, n$  all distinct. Then

$$i) \quad \lim_{t \rightarrow \infty} a_t = a = [a_1, a_2, \dots, a_n]', \text{ with } a_i = (1 - u_i^2) \prod_{j \neq i} \frac{1 - u_i u_j}{1 - \frac{u_i}{u_j}}$$

$$ii) \quad \lim_{t \rightarrow \infty} \dot{Y}_t = \dot{Y} = V / \prod_{i=1}^n u_i^2$$

$$iii) \quad \lim_{t \rightarrow \infty} C_t = C = \{c_{ij}\}, c_{ij} = V(1 - u_i u_j) \prod_{k \neq j} \frac{1 - u_k u_i}{u_k - u_i} \prod_{h \neq i} \frac{1 - u_h u_j}{u_h - u_j}$$

$$iv) \quad \lim_{t \rightarrow \infty} W_t = W = B^{-1} G C G' B'^{-1} - G C G' = \{w_{ij}\},$$

$$w_{ij} = (1 - (\beta_i \beta_j)^{1/2}) \frac{c_{ij}}{u_i u_j} \quad i, j = 1, 2, \dots, n;$$

#### PROOF

From Theorem 6.2 or 6.3,  $\lim_{t \rightarrow \infty} \{C, R, \dot{Y}, a\}_t = \{C, R, \dot{Y}, a\}$  all exist, unique,

and  $C, R$  are non singular. Moreover,

$$\{c_{ij}\} = C = (I - af)R \quad (6.5)$$

$$Q = B'^{-1} G'^{-1} Q G^{-1} B^{-1} + f' f, \quad Q = C^{-1} V = \{q_{ij}\} \quad (6.6)$$

$$\mathbf{a} = \mathbf{a}(n) = \mathbf{Q}^{-1} \mathbf{f}, \quad (6.7)$$

where  $\mathbf{a}'(n) = [a_1(n), a_2(n), \dots, a_n(n)]$ .

From (6.6),

$$q_{ij} = \frac{1}{1 - u_i u_j} \quad (6.8)$$

Now, multiplying (6.7) by  $\mathbf{Q}$  gives

$$\mathbf{Q} \mathbf{a}(n) = \mathbf{f}. \quad (6.9)$$

For  $n=1$ , (6.8) and (6.9) gives

$$a_1(1) = 1 - u_1^2 \quad (6.10)$$

For  $n \geq 2$ ,  $i=1, 2, \dots, n$ ; from (6.9), we have

$$\sum_{j=1}^n q_{ij} a_j(n) = 1; \quad i = 1, 2, \dots, n \quad (6.11)$$

Therefor,

$$a_n(n) = (1 - \sum_{j=1}^{n-1} q_{nj} a_j(n)) / q_{nn}$$

substituting for  $a_n(n)$  in the first  $n-1$  equations of (6.11),

$$\sum_{j=1}^{n-1} \frac{q_{nn} q_{ij} - q_{nj} q_{in}}{q_{nn} - q_{in}} a_j(n) = 1$$

This gives

$$\sum_{j=1}^{n-1} \frac{1 - \frac{u_j}{u_n}}{1 - u_j u_n} q_{ij} a_j(n) = 1 \quad (6.12)$$

Since  $\mathbf{a}(n)$  is unique and (6.11) is true for all  $n$ ,

$$a_j(n) = \frac{1 - u_j u_n}{1 - \frac{u_j}{u_n}} a_j(n-1); \quad j = 1, 2, \dots, n-1$$

This , together with (6.10) proves (i).

ii)

From (6.4)

$$V = \dot{Y} - f R f'$$

$$= (1 - f a) \dot{Y} = (1 - \sum_{i=1}^n a_i) \dot{Y}$$

$$\text{but from (i) it follows that } \sum_{i=1}^n a_i = 1 - \prod_{i=1}^n u_i^2$$

iii)

From (6.5) we have

$$\begin{aligned} c_{ij} &= (c_{ij}/u_i u_j) - a_i a_j \dot{Y} \\ &= u_i u_j a_i a_j \dot{Y} / (1 - u_i u_j) \end{aligned}$$

the result follows from (i) and (ii)

iv)

Easily derived from the definition of  $W$

#### COROLLARY 6.4.1.

If  $\beta_k = \beta$  for all  $k$  then (i) reduces to the EWR result of Dobbie (1963).

The theorem is of practical interest mainly for periodic models with distinct complex eigenvalues lying inside or on the unit circle. For a real observation series a similar model with real  $G$  would be adopted and the corresponding limiting values are easily derived. For example, if  $G = \begin{bmatrix} \cos w & \sin w \\ -\sin w & \cos w \end{bmatrix}$ , an alternative NDBM can be considered with  $G = \begin{bmatrix} e^{iw} & 0 \\ 0 & e^{-iw} \end{bmatrix}$ .

A more general procedure for finding the adaptive coefficients can be deduced using the following

### THEOREM 6.5.

Let  $\{f, G, V, H\}$  be a constant NWBM with  $\lim_{t \rightarrow \infty} C_t = C$ , non singular, and  $A = CF'V^{-1}$ . Then, the two transformations  $(I - AF)H$  and  $H^{-1}$  have identical characteristic polynomials.

If  $H = \beta I$ , then  $\beta^{-1}(I - AF)G$  and  $G^{-1}$ , and  $(I - AF)G$  and  $\beta G^{-1}$  have identical characteristic polynomials.

### PROOF :

Since  $\lim_{t \rightarrow \infty} C_t = C$  is non singular, from the NWBM properties, we have

$$C = (I - AF)R \quad , \quad R = HCH'$$

This gives

$$CH'^{-1}C^{-1} = (I - AF)H \tag{6.13}$$

The result follows from (6.13).

The above results can be used to calculate the limiting adaptive coefficients for any observable NDBM if its state covariance matrix converges to a non singular limit. In particular, for  $\{f, G, V, \beta I\}$  NDBM's such that  $0 < \beta/\lambda_i^2 < 1$ , where  $\lambda_1, \lambda_2, \dots, \lambda_n$  are the eigenvalues of  $G$  as in the following NDBM's that are commonly used in practice apart from the one given in Theorem 6.4.

### COROLLARY 6.5.1.

If  $f = [1, 0, 0, \dots, 0]$  and  $G = J(\lambda)$  where,  $J(\lambda)$  has a Jordan form and  $0 < \beta < \lambda^2$ .

Then

$$\lambda a_i + a_{i+1} = \binom{n}{i} \rho^i, \quad i = 1, 2, \dots, n$$

where  $\rho = \frac{(\lambda^2 - \beta)}{\lambda}$ ,  $a' = [a_1, a_2, \dots, a_n]$  and  $a_{n+1} = 0$ .



## PROOF

It is easily seen that the above NDBM is observable. Since  $0 < \beta < \lambda^2$ , using Theorem (6.2),  $\lim_{t \rightarrow \infty} \{C, a\}_t = \{C, a\}$  both uniquely exist. Moreover  $C$  is non singular.

From Theorem (6.5)

$$\det(\beta G^{-1} - xI) = \det((I - af)G - xI), \quad H = \beta^{-1}G \quad \text{for all } x.$$

This gives

$$\det \begin{bmatrix} \alpha - r_1 & 1 & 0 & 0 & 0 \\ -r_2 & \alpha & 1 & . & . \\ . & . & \alpha & . & . \\ . & . & . & . & . \\ . & . & . & . & 1 \\ -r_n & 0 & 0 & 0 & \alpha \end{bmatrix} = (\beta/\lambda - x)^n$$

where,  $\alpha = \lambda - x$  and  $r_i = \lambda a_i + a_{i+1}$ ,  $a_{n+1} = 0, i = 1, 2, \dots, n$ . Hence,

$$\sum_{i=0}^n r_i \alpha^{n-i} = (\rho - \alpha)^n \quad (6.14)$$

The result follows from the comparison of the coefficients of each power of  $\alpha$  in this equation.

## COROLLARY 6.5.2.

IF  $f = [1, 0, 0, \dots, 0]$  and  $G$  is an upper triangular matrix with entries 1 and  $0 < \beta < 1$ , then

$$a_i = \sum_{j=i}^n \binom{n}{j} (1-\beta)^j \beta^{n-j}; \quad i = 1, 2, \dots, n.$$

## PROOF

Following the steps as in the proof of Corollary 6.5.1, the alternative form of (6.14) in the variable  $x$  is

$$(1-x)^n - a_1(1-x)^{n-1} + a_2x(1-x)^{n-2} + \dots + (-1)^n a_n x^{n-1} = (\beta-x)^n \quad (6.15)$$

Writing this in powers of  $(1-x)$  and collecting the terms in the coefficient of  $(1-x)^i$  from both sides of (6.15), gives

$$\sum_{k=0}^{n-i} \binom{i+k-1}{k} a_{i+k} = \binom{n}{i} (1-\beta)^i, \quad i=1,2,\dots,n.$$

The values of  $a_i$  can be found successively from the above equation.

### COROLLARY 6.5.3.

In Corollary 6.5.2, if  $f$  is replaced by  $f = 1.1.1.\dots.1$  then

$$a_i = \binom{n}{i} (1-\beta)^i \beta^{n-i} \quad ; i=1,2,\dots,n$$

### PROOF

Similar calculations give the alternative form of (6.15) as

$$(1-x)^n (1 - \sum_{i=1}^n a_i) + \sum_{k=1}^n (-1)^k x^k (1-x)^{n-k} a_k = (\beta-x)^n$$

The result follows from comparison with the terms of

$$(\beta-x)^n = [\beta(1-x) - x(1-\beta)]^n = \sum_{k=0}^n (-1)^k \binom{n}{k} x^k (1-\beta)^k \beta^{n-k} (1-x)^{n-k}$$

### COROLLARY 6.5.4.

If  $f = [1,0,0,\dots,0]$ ,  $G = \begin{bmatrix} 0 & I \\ -1 & 0 \end{bmatrix}$  with  $0 < \beta < 1$ , then  $a_1 = 1 + \beta^n$  and  $a_i = 0$  ; otherwise.

### PROOF

Similar calculations shows that the alternative form of (6.15) is

$$x^n + a_2 x^{n-1} + \dots + a_n x + a_1 - 1 = x^n + \beta^n$$

The result follows.

Now, given that  $\lambda_1, \lambda_2, \dots, \lambda_n$  ( not necessary distinct ) are the eigenvalues of  $G$  for a constant NDBM, the restrictions on  $B$  ensures the existence of  $\lim_{t \rightarrow \infty} C_t = C$  as a proper covariance matrix. Denoting the eigenvalues of  $CR^{-1}G$  or equivalently  $(I - af)G$  by  $\rho_i$  ;  $i = 1, 2, \dots, n$  we have the following

### THEOREM 6.6.

Given any constant observable NWBM  $\{f, G, V, H\}$ , for which  $\lim_{t \rightarrow \infty} C_t = C$  is positive definite,

$$\lim_{t \rightarrow \infty} \left( \prod_{i=1}^n (1 - \lambda_i B) y_t - \prod_{i=1}^n (1 - \rho_i B) e_t \right) = 0$$

where  $B$  is the backward shift operator and  $e_t$  is the one step ahead forecast error.

### PROOF.

Since  $\lim_{t \rightarrow \infty} C_t = C$  is positive definite,  $\lim_{t \rightarrow \infty} \{R, a\}_t = \{R, a\}$  exists and  $R$  is non singular.

Let  $\rho_i, i=1, 2, \dots, n$  be the eigenvalues of  $X = CR^{-1}G = (I - af)G$ .

A direct application of the Bayes theorem in updating (4.3) , using (4.1) with univariate observations, gives

as  $t \rightarrow \infty$ ,

$$m_t = Gm_{t-1} + ae_t \tag{6.16}$$

$$G^{-1}m_t = R^{-1}Gm_{t-1} + V^{-1}f'y_t, \text{ or}$$

$$m_t = Xm_{t-1} + ay_t \tag{6.17}$$

From (6.16) and the identity  $e_{t+1} = y_{t+1} - fGm_t$ , we have

$$e_{t+1} = y_{t+1} - fG(I - BG)^{-1}ae_t$$

Hence,

$$(1 + BfG(I - BG)^{-1}a)e_{t+1} = y_{t+1} \quad (6.18)$$

The same identity with (6.17) gives,

$$e_{t+1} = y_{t+1} - fG(I - BX)^{-1}ay_t, \text{ or}$$

$$e_{t+1} = (1 - BfG(I - BX)^{-1}a)y_{t+1} \quad (6.19)$$

Note that,  $\det(I - BG)$  and  $\det(I - BX)$  are  $\prod_{i=1}^n (1 - \lambda_i B)$  and  $\prod_{i=1}^n (1 - \rho_i B)$  respectively.

(6.18) and (6.19) can be rewritten as

$$P_1(B)e_{t+1} = \prod_{i=1}^n (1 - \lambda_i B)y_{t+1} \quad (6.20)$$

$$\prod_{i=1}^n (1 - \rho_i B)e_{t+1} = P_2(B)y_{t+1} \quad (6.21)$$

Where  $P_1(B)$  and  $P_2(B)$  are polynomials of degree  $n$  in  $B$ .

From (6.20) and (6.21),

$$\prod_{i=1}^n (1 - \lambda_i B) \prod_{i=1}^n (1 - \rho_i B) = P_1(B)P_2(B)$$

The result follows using the factorisation theorem.

This result includes the EWR result of McKenzie (1976) and hence the same result obtained by Godolphin and Harrison (1975) through special DLM formulations. These are obtained for scalar discount matrices. That is  $B = \beta I$ . The Normality assumption can be relaxed since (2.16) and (2.17) can also be obtained using minimum variance unbiased linear estimation. Hence the results may be extended beyond the Normal models.

#### 6.4. A GENERAL LIMITING THEOREM :

Any observable constant NWBM  $\{f, G, V, B\}$  is similar to the canonical constant NWBM  $M = \{f, G, V, H\}$  where , writting  $B = \text{diag}\{\beta_1, \beta_2, \dots, \beta_n\}$  ,

$f = [1, 0, 0, \dots, 0]$  and  $H = [h_{ij}]$  such that

$$h_{i,i} = -h_{i,i+1} = \lambda_i / (\beta_i)^2 = u_i, \text{ if } i = 1, 2, \dots, n$$

$h_{i,j} = 0$  otherwise, and  $0 < \beta_i^2 / \lambda_i < 1$ ,  $H^{-1} = [a_{ij}]$  with  $a_{ij} = \frac{1}{u_j}$  for  $j \geq i$  and

$a_{ij} = 0$  otherwise.

It follows from Theorem 6.3, that  $\lim_{t \rightarrow \infty} C_t = C = Q^{-1}V$  exists where the precision recursion can be rearranged to give the Liapunov equation

$$H'Q = QH^{-1} + H'f'f$$

This allows an easy sequential term by term evaluation of  $Q = \{q_{ij}\} :-$

$$q_{11} = \frac{u_1^2}{(u_1^2 - 1)} \quad ; \quad q_{12} = \frac{q_{11}}{(u_1 u_2 - 1)}$$

$$q_{1i} = u_1 u_{i-1} \frac{q_{1,i-1}}{(u_i u_1 - 1)} \quad \text{for } i > 2$$

and

$$q_{i,k} = u_{i-1} \frac{q_{i-1,k}}{u_i} + \frac{S_{i,k}}{u_k}$$

where

$$S_{i,k} = q_{i,k} + S_{i,k-1} = \sum_{j=1}^k q_{ij} \quad \text{and} \quad q_{ij} = q_{ji}.$$

It follows that

$$i) \quad C = Q^{-1}V$$

$$ii) \quad a = Q^{-1}f', \quad a_1 = 1 - \prod_{i=1}^n u_i^{-2} \quad \text{and} \quad a_n = (-1)^{n+1} (1 - a_1) \prod_{i=1}^n (u_i u_n - 1)$$

$$iii) \quad \dot{Y} = \frac{V}{(1 - a_1)} = V \prod_{i=1}^n u_i^2$$



$$iv) \quad W = HCH' - CGG'$$

### 6.5. RELATIONS WITH ARIMA MODELS :

Let  $Y_t$  be a random time series generated according to an ARIMA model

$$\prod_{i=1}^n (1 - \lambda_i B) Y_t = \prod_{i=1}^n (1 - \rho_i B) a_t$$

where  $0 < |\lambda_i| \leq 1$ ,  $0 < |\rho_i| < 1$ ,  $i = 1, 2, \dots, n$  and  $a_t$  is such that  $E a_t = 0$ ,  $E a_t^2 = \sigma^2$  and  $E a_t a_{t+k} = 0$  for all  $k > 0$ . The appropriate Box-Jenkins (1970) predictor replaces  $a_t$  by the one step ahead prediction error  $e'_t$  and it is well known that  $\lim_{t \rightarrow \infty} e'_t = a_t$ . Applying the appropriate Dynamic Model  $\{f, G, V, \dots\}$  to the realised series  $y_t$ ,

$$\lim_{t \rightarrow \infty} \left\{ \prod_{i=1}^n (1 - \lambda_i B) y_t - \prod_{i=1}^n (1 - \rho_i B) e_t \right\} = 0$$

Hence  $\lim_{t \rightarrow \infty} |e_t - e'_t| = 0$  and with probability one, the limiting Box-Jenkins forecast function is equivalent to that of the Dynamic Model.

For an unbalanced ARIMA process

$$\prod_{i=1}^p (1 - \lambda_i B) y_t = \prod_{i=1}^q (1 - \rho_i B) a_t$$

Let  $n = \max \{ p, q \}$ . Then given any  $\epsilon > 0$ , if  $n = p$  ( or  $n = q$  ) by taking  $p - q$  of the  $\rho_i$ 's ( or  $q - p$  of the  $\lambda_i$ 's ) approximately close to zero,  $\lim_{t \rightarrow \infty} |e_t - e'_t| < \epsilon$  is assumed.

Thus, in the sense of limiting forecast functions, all ARIMA processes can be modelled by constant NWBM's. In fact if the limiting posterior state variance is taken as the original prior variance then, the forecast functions can be identical to that of ARIMA models all the way through the sequential analysis. However, as stated earlier, the NWBM provides parameter informations in a sensible way. This simplifies explaining and controlling the process and models behaviour.

## 6.6. SUMMARY :

This chapter is concerned with the derivation of some interesting limiting results regarding the posterior covariance matrices, the adaptive coefficients and the parsimonious transfer functions using some well known and simple canonical representations. In particular a simple transformation procedure within similar models is discussed in 6.2. Limiting results regarding the forecasting variance, the adaptive vector, precision and covariance matrices are given in 6.3 and 6.4. The link with Box-Jenkins ARIMA models in terms of forecast functions is discussed in 6.5.

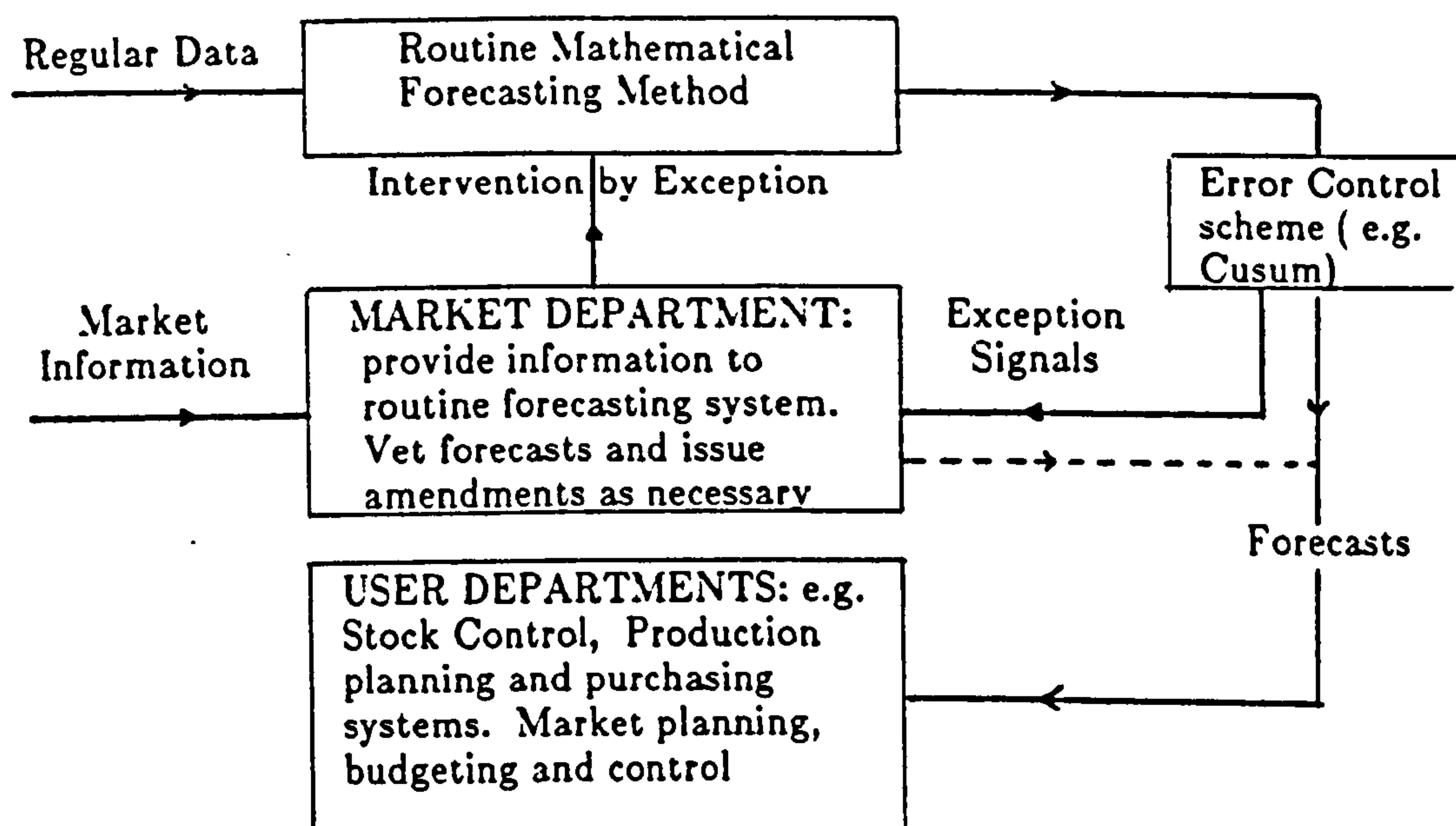
## CHAPTER SEVEN

### MULTIPROCESS MODELS WITH CUSUMS

#### 7.1. INTRODUCTION :

Many analyses of statistical data sets are based on the assumptions that the input data is free from exceptions, properly collected and well behaved. However, in practice, it is hard to believe that all these smoothness properties can be guaranteed. Often the data contains missing values, outliers and sudden structural changes in the process behaviour. It is then believed that the occurrence of any of these events in sequential procedures causes model breakdown and damages the available prior information as pointed out by Jaynes(1983). These events call for model revision and amendments. In forecasting, the principle of ' Management by Exception ' is widely applied. This constitutes mathematical methods producing routine forecasts required by various decision makers. These forecasts are acted upon unless exceptional circumstances occur due either to the anticipation of a major change arising from the use of reliable market information, (see Harrison and Scott(1965) and Harrison (1967)) or, to the occurrence of some unforeseen change in the pattern of demand which causes unusual forecasting errors and consequently a model breakdown. A flowchart of the principle is given in Fig.7.2.

A Management by Exception Forecasting System (Fig.2)





In this chapter, efficient statistical models are introduced to deal automatically with exceptions. Ameen and Harrison (1983 c). Section 2 reviews the historical background and developments. The backward Cumulative Sum (CUSUM) Statistic is reviewed in Section 3. The Multiprocess model approach of Harrison and Stevens is reviewed in the light of discounting in Section 4. In Section 5, the ideas from the backward CUSUM and the multiprocess models of Harrison and Stevens together with the Modified NDBM's are combined to provide both economical and efficient multiprocess models called Multiprocess Models with CUSUM's. These eliminate many unnecessary computations involved in the existing multiprocess models and protect prior information on components unchanged structurally when changes occur in other components, Ameen(1983 a).

## 7.2. HISTORICAL BACKGROUND AND DEVELOPMENTS :

Woodward and Goldsmith (1964) have employed Backward CUSUM tests to detect unanticipated changes in demand. The procedure is given by Page (1954), Barnard(1959), Ewan and Kemp(1960) and Ewan(1963). Harrison and Davies(1964) used CUSUM's for controlling routine forecasts of product demand and provided simple recursion formulas to reduce data storage problems. These are reviewed in Section 3. More details on the CUSUM statistic can be found in Van Dobben De Bruyn (1968) and Bissell(1969). (For general sequential tests see Wald(1947)).

Previously having detected a change, ad hoc intervention procedures were applied. The first routine computer forecasting systems for stock control and production planning, employed Exponential Weighted Moving Averages ( EWMA ) and Holts linear growth model, with or without seasonal components. All the forecasting methods used limiting predictors which assume a reasonably long history of well behaved data. The occurrence of a major change means that, in some respects, the current data does not reflect a well behaved process and that there is greater uncertainty than usual about the future. Hence the next data points will be very informative in removing much of this increased uncertainty and should be given more weight than they would be allocated by the limiting



predictor. For example, consider the use of EWMA with forecast function

$$F_t(k) = m_t \quad \text{where ,}$$

$$m_t = m_{t-1} + a_t e_t \quad . \quad \text{and}$$

$$e_t = y_t - m_{t-1} \quad , \quad a_t = 0.2$$

This may be written as

$$m_t = 0.8 m_{t-1} - 0.2 y_t$$

where  $y_t$  is the observed demand for period  $t$ .

Suppose that in the limiting case, the variance  $V(e_t) = \text{Var} ( Y_t | D_{t-1} ) = 125$  and that  $m_t = 100$  with an associated variance of 20. As a result of a CUSUM signal, Marketing department may wish to intervene by stating that their best estimate of the level is not now 100 but 150 and that their variance associated with this estimate is not 25 but 300. In the past there was no formal way of dealing with this. Classical time series methods based upon the assumptions of derived stationarity are inappropriate. Typically what was done was to introduce a change in the adaptive coefficient  $a$  which here is originally 0.2. One procedure put

$$a_{t+i} = \begin{cases} 0.9 - i/10 & \text{if } i=1,2,..6 \\ 0.2 & \text{if } i > 6 \end{cases}$$

This approach is not very satisfactory nor does it generalise well in dealing with other kinds of change.

The DLM's of Harrison and Stevens(1971,1976) and the NWBM's introduced in Chapter 4 provide a formal way of combining subjective judgements and data. In the above example the adopted DLM is

$$Y_t = \theta_t + v_t \quad ; \quad v_t \sim N[0;100]$$

$$\theta_t = \theta_{t-1} + w_t \quad ; \quad w_t \sim N[\overline{w}_t; W_t]$$

$\theta_t$  represents the underlying market level at time  $t$ ,  $W_t = 5$  and usually  $\bar{w}_t = 0$ . In the limit  $(\theta_t|D_t) \sim N[m_t; 20]$ . The limiting recurrence relationship is  $m_t = m_{t-1} + 0.2e_t$  and the limiting one step ahead forecast distribution is  $(Y_{t+1}|D_t) \sim N[m_t; 125]$ . Hence the EWMA provides the limiting point forecasts. In the example  $(\theta_t|D_t) \sim N[100; 20]$ , the market information is communicated as  $w_{t+1} \sim N[50; 280]$ . Now the one step ahead forecast distribution is not  $(Y_{t+1}|D_t) \sim N[100; 125]$  but  $(Y_{t+1}|D_t) \sim N[150; 400]$ . Immediately the recursive equation departs from its limit and becomes  $m_{t+1} = m_t - 50 + 0.75e_{t+1}$ . Provided future interventions do not occur the adaptive coefficient  $a_{t+1}$  returns fairly quickly to its limiting value of 0.2. Note that the same results can be achieved using an NDBM with discount factor  $\beta = 0.8$  where at the intervention time its value is reduced to  $\beta^* = 0.066$  with adjusting the state prior mean from 100 to 150.

Other related works are those of Kalman (1963), Smith (1979) and Harrison and Akram (1983).

Bayesian forecasting provides a means of dealing with specified types of major changes. These forms of change are modelled so that the forecasting system deals with them in a prescribed way. The initial implementation of the resulting multiprocess models is described in Harrison and Stevens (1971, 1975, 1976). These are reviewed in Section 4. In addition to the limitations and drawbacks of single DLM's mentioned in Chapter 3, they involve unnecessary computations. Smith and West (1983) have applied these models to monitoring Kidney transplants. Restricting the models to steady state processes, these are generalised to non Normal models by Souza (1981). Limited success is achieved by Gathercole and Smith (1983) in reducing the computation efforts by removing redundant models according to some pre specified rules. For another attempt see Makov (1983). In general practice the development and existence of these methods replaced the control chart techniques.

### 7.3. THE BACKWARD CUSUM :

Control charts provide simple and effective tools for detecting changes and departures from specific target values and are particularly valuable in quality control. Page(1954) used Cumulative Sum charts in detecting changes in process level. In fact they can be used for detecting the amount and direction of these changes. Further developments on this topic and the use of Average Run Length to increase the sensitivity of the signals can be found in Barnard(1959), Woodward and Goldsmith(1964) and Ewan and Kemp(1960). Given  $y_t$  as the observed process value and  $T$  as a target value, the CUSUM statistic  $S_t$  is defined for each time  $t$  as  $S_t = S_{t-1} + e_t$ , where  $e_t = y_t - T$ .

Since  $(e_t | D_t) \sim N[0; \hat{Y}_t]$ ,  $S_t$  is Normal with zero mean. Then choosing two positive constants  $L_0$  and  $\alpha$ , visual inspection can be carried out with the graph of  $(t, S_t)$  and using a V-shaped hole cut out from a piece of cardboard and placed on the graph with the vertex of the V-mask pointed horizontally with a distance  $(L_0/\alpha)+1$  from the leading point. The edges of the V-mask being apart with angle  $2\psi$ , where  $\tan\psi = \alpha$ . No change is signaled as long as the CUSUM curve remains inside the V-mask. Harrison and Davies (1964) developed the method for monitoring forecasts of product demand. The target value is the one step ahead point forecasts so that the  $e_t$  series is that of the one step ahead forecast errors. In order to reduce computer storage, a conventionally simple and economical algorithm was employed.

Define

$$\bar{d}_{t+1} = \min\{L_0, \bar{d}_t\} + \alpha - e_{t+1}/\hat{Y}_{t+1}, \quad \text{and} \quad (7.1)$$

$$d_{t+1} = \min\{L_0, d_t\} + \alpha + e_{t+1}/\hat{Y}_{t+1} \quad (7.2)$$

where  $\hat{Y}_{t+1}$  is the one step ahead forecast variance. A change is signaled if and only if  $\min\{\bar{d}_t, d_t\} < 0$ . Initially  $\bar{d}_0 = d_0 = L_0$ . In choosing  $L_0$  and  $\alpha$ , the following facts may be used as guide lines.

### THEOREM 7.1.

Given that the V-mask has not signaled a change at time  $t$ , for time  $t+1$ ,



i) A change will be signaled, if

$$|e_{t+1}/\hat{Y}_{t+1}| > L_0 + \alpha \quad (7.3)$$

ii) Change will not be signaled, if

$$|e_{t+1}/\hat{Y}_{t+1}| \leq \alpha \quad (7.4)$$

### PROOF

i) From (7.1)

$$\bar{d}_{t+1} = \min\{L_0 + \alpha - e_{t+1}/\hat{Y}_{t+1}, \bar{d}_t - \alpha - e_{t+1}/\hat{Y}_{t+1}\}$$

Hence, given (7.3),  $\bar{d}_{t+1} < 0$ .

ii) Substituting (7.4) in each of (7.1) and (7.2), it follows that  $\bar{d}_{t+1} > 0$  and  $\underline{d}_{t+1} > 0$ .

### 7.4. NORMAL WEIGHTED BAYESIAN MULTIPROCESS MODELS :

The DLM - multiprocess models as proposed by Harrison and Stevens (1971, 1975, 1976) are reviewed in the light of the NWBM's introduced in Chapter 3.

The set  $\{(M^{(1)}, \mathbf{p}^{(1)}), \dots, (M^{(N)}, \mathbf{p}^{(N)})\}$  is a multiprocess NWBM with N model components such that : for  $i = 1, 2, \dots, N$  ;  $M^{(i)} = \{\mathbf{F}^{(i)}, \mathbf{G}^{(i)}, V^{(i)}, \mathbf{H}^{(i)}\}_t$  represents a NWBM,

$$\mathbf{p}^{(i)} = \{P^{(i)}, \boldsymbol{\pi}^{(i)}\}_t, \text{ where}$$

$P_t^{(i)} \geq 0$  is the posterior probability of model i at time t,  $\sum_{i=1}^N P_t^{(i)} = 1$ ,  $\boldsymbol{\pi}_t^{(i)} = [\pi_{1t}^{(i)}, \dots, \pi_{Nt}^{(i)}]$  is

the model transition probability vector such that  $\pi_{kt}^{(i)} = P[M_t^{(k)} | M_{t-1}^{(i)}]$ , that is the

probability that at time t model k operates given that the operational model at time t-1

was  $M^{(i)}$ . Initially assume that the  $\mathbf{p}^{(i)}$ 's and the  $M^{(i)}$ 's are known, although in practice

the  $V_t^{(i)}$ 's are estimated on-line. At time t-1, let there be N conditional posterior

distributions  $(\mathbf{0}_{t-1} | M_{t-1}^{(i)}, D_{t-1}) \sim N[\boldsymbol{\mu}_{t-1}^{(i)}; \mathbf{C}_{t-1}^{(i)}]$ . The  $N^2$  conditional one step ahead

forecasts are :

$$(Y_t | M_t^{(j)}, M_{t-1}^{(i)}, D_{t-1}) \sim N[\hat{\mathbf{y}}_t^{(ij)}; \hat{\mathbf{Y}}_t^{(ij)}]$$

where

$$\hat{\mathbf{y}}_t^{(ij)} = \mathbf{F}^{(j)} \mathbf{G}^{(j)} \mathbf{m}_{t-1}^{(i)} \quad \text{and} \quad \hat{\mathbf{Y}}_t^{(ij)} = \mathbf{F}^{(j)} \mathbf{R}_t^{(ij)} \mathbf{F}^{(j)'} + \mathbf{V}_t^{(j)}, \text{ with}$$

$$\mathbf{R}_t^{(ij)} = \mathbf{H}_t^{(j)} \mathbf{C}_{t-1}^{(i)} \mathbf{H}_t^{(j)'}$$

The one step ahead forecast distribution is expressed as a mixture of  $N^2$  Normals

$$(\mathbf{Y}_t | D_{t-1}) \sim \sum_{j=1}^N \sum_{i=1}^N \pi_j^{(i)} P_{t-1}^{(i)} N(\hat{\mathbf{y}}_t^{(ij)}; \hat{\mathbf{Y}}_t^{(ij)})$$

Also, given  $N$  posterior models at time  $t-1$ ,  $N^2$  prior models are produced for time  $t$ , for which given the data at time  $t$  and using Bayes theorem, the  $N^2$  posterior models for time  $t$  are :

$$(\boldsymbol{\theta}_t | M_t^{(j)}, M_{t-1}^{(i)}, D_t) \sim N(\mathbf{m}_t^{(ij)}; \mathbf{C}_t^{(ij)}) \quad ; \quad i, j = 1, 2, \dots, N$$

where

$$\mathbf{m}_t^{(ij)} = \mathbf{G}^{(j)} \mathbf{m}_{t-1}^{(i)} + \mathbf{A}_t^{(ij)} \mathbf{e}_t^{(ij)},$$

$$\mathbf{C}_t^{(ij)} = (\mathbf{I} - \mathbf{A}_t^{(ij)} \mathbf{F}^{(j)}) \mathbf{R}_t^{(ij)},$$

$$\mathbf{A}_t^{(ij)} = \mathbf{R}_t^{(ij)} \mathbf{F}^{(j)'} (\hat{\mathbf{Y}}_t^{(ij)})^{-1}$$

$$\mathbf{e}_t^{(ij)} = \mathbf{y}_t - \hat{\mathbf{y}}_t^{(ij)},$$

and given the likelihood :

$$L(M_t^{(j)} | M_{t-1}^{(i)}, D_t) \propto |\hat{\mathbf{Y}}_t^{(ij)}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{e}_t^{(ij)'} (\hat{\mathbf{Y}}_t^{(ij)})^{-1} \mathbf{e}_t^{(ij)} \right\}$$

the associated  $N^2$  posterior probabilities are :

$$P_t^{(ij)} \propto L(M_t^{(j)} | M_{t-1}^{(i)}, D_{t-1}) \pi_j^{(i)} P_{t-1}^{(j)}$$

In practice, in order to keep the computations manageable, the same collapsing procedure as defined by Harrison and Stevens is used to complete the cycle. calculating

$$\{(\boldsymbol{\theta}_t | M_t^{(j)}, D_t), P_t^{(j)}; j = 1, 2, \dots, N\} \text{ as}$$



$$P_t^{(j)} = \sum_{i=1}^N P_t^{(ij)} ,$$

$$m_t^{(j)} = \sum_{i=1}^N P_t^{(ij)} m_t^{(ij)} / P_t^{(j)} ,$$

$$C_t^{(j)} = \sum_{i=1}^N P_t^{(ij)} [C_t^{(ij)} + (m_t^{(ij)} - m_t^{(j)})(m_t^{(ij)} - m_t^{(j)})' / P_t^{(j)}] .$$

As in Harrison and Stevens(1976) models, the NWB multiprocess models are partitioned into Class I models, in which there is no transition between models, that is  $\pi_{ik}^{(j)}$  has all elements but  $\pi_{ii}^{(j)}$  equal to zero, and Class II models, in which such transitions exists and models operate interactively. The former class is used on-line for model discrimination, model estimation and hypothesis tests. Class II models are used for modelling some prescribed types of disturbances and alternative models. As explained in Chapter 4, all the normal DLM's can be formulated as NWBM's. In this sense, the multiprocess normal DLM applications can be counted as multiprocess NWBM applications. The former has worked well in analysing processes with disturbances such as outliers and sharp changes in level and seasonality components. However, in many cases slope changes have not been modelled as successfully as would be desired. Brown(1983) also commented on this problem. This is largely because such changes are generally small compared to both the random variation and to changes in process level. Hence slope changes are sometimes identified as a series of level changes. Smith and Cook(1980), Harrison and Davies (1964) also commented on the difficulty in distinguishing level and slope changes in CUSUM analysis. A further criticism of these multiprocess models is that they involve unnecessary computations. These problems are overcome by introducing a new class of multiprocess models using a combination of Class I and Class II models with the Backward CUSUM statistic as a control device for shifting model operations from Class I to Class II models. Class I is retained when one of the members of Class II attains some prespecified threshold probability limit.

## 7.5. MULTIPROCESS MODELS WITH CUSUMS :

In most multiprocess Class II applications, there is a preferred model  $M^{(1)}$  called the 'mother' model by Gathercole and Smith (1983). The analogy with quality control is that  $M^{(1)}$  describes the data as long as it is behaving in an expected way. The other models  $M^{(i)}; i \geq 2$ , generally model some significant type of departure from the norm. In particular outliers or mavericks and significant changes in the trend are often modelled.

In the new approach the mother model  $M^{(1)}$  is represented by a NWBM  $\{f, G, V, H\}_t$ . This model produces forecasts which are used unless a departure from normal is signaled by the CUSUM scheme which operates on the one-step-ahead forecast errors. Then, starting with the latest observation which helped to trigger the signal, the other models are applied. All the models then operate in a multiprocess Class II way, with a high probability of transition to the 'mother' model ( $M^{(1)}$ ), until such a time that the posterior probability  $P_t^{(1)}$  of model  $M^{(1)}$  exceeds a given value. When this happens, model  $M^{(1)}$  begins to operate alone and the CUSUM scheme is reset. When one model is operating, although all forecasts are based upon model  $M^{(1)}$ , the competing models are being prepared in readiness for the multiprocess phase.

For example, consider  $G = \text{diag}\{G_1, G_2\}$  in which  $G_1$  represents a trend component and  $G_2$  a seasonal component. Let  $M^{(2)}$  model major changes in trend,  $M^{(3)}$  model outliers, and model  $M^{(4)}$  model major changes in seasonality.

Let  $\theta_1$  and  $\theta_2$ , be the trend and seasonal parameter vectors respectively. For  $M^{(i)}$ ,  $i=1,2,3,4$ , let the posterior state distributions at time  $t$ , be given as

$$\left( \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \right)_t | D_t, M_t^{(i)} \sim N \left[ \begin{bmatrix} m_1^{(i)} \\ m_2^{(i)} \end{bmatrix}; \begin{bmatrix} G_1^{(i)} & C_3^{(i)} \\ C_3'^{(i)} & G_2^{(i)} \end{bmatrix} \right]$$

The priors for time  $t+1$  are then formed from the posteriors as follows:

$$\left( \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \right)_{t+1} | D_t, M_t^{(i)} \sim N \left[ \begin{bmatrix} G_1 m_1^{(i)} \\ G_2 m_2^{(i)} \end{bmatrix}; \begin{bmatrix} R_1^{(i)} & R_3^{(i)} \\ R_3'^{(i)} & R_2^{(i)} \end{bmatrix} \right]$$

where

$$R_k^{(i)} = G_k C_k^{(i)} G_k' / \beta_k^{(i)} ; \quad k=1,2 \text{ ; and}$$

$$l = \begin{cases} 2 & \text{if } k=1 \text{ and } i=2 \\ 4 & \text{if } k=2 \text{ and } i=4 \\ 1 & \text{otherwise} \end{cases}$$

$$R_3^{(1)} = G_1 C_3^{(1)} G_2' / (\beta_1^{(1)} \beta_2^{(1)}) ,$$

$$0 < \beta_1^{(2)} \ll \beta_1^{(1)} = \beta_1^{(3)} = \beta_1^{(4)} < 1 ,$$

$$0 < \beta_2^{(4)} \ll \beta_2^{(1)} = \beta_2^{(2)} = \beta_2^{(3)} < 1$$

The general principle is to derive the marginal prior for the parameter block characterising the change, from the posterior but to take all other information from the mother model. This aims at keeping the information on other components as stable as possible in order to prepare good estimates for the changes. Otherwise the covariance structure between the model components might produce violent fluctuations in the estimates of the presumed stable components. In addition a set of preparatory model probabilities is calculated using Bayes theorem, so that

$$P_t^{(i)} \propto L(y_t | M^{(i)}) P_{t-1}^{(i)}$$

When the CUSUM signals a change all these preparatory values are used as starting values for the NWB multiprocess phase. However, generally all information other than that marginal characterising the change continues to be taken from the mother model. In order to exercise control over the response of models to exceptional events, a guard procedure on the observation variance for alternative models is used. This is performed by choosing  $V^{(3)}$  and  $\beta^{(i)} ; i > 1$ , so that the one-step-ahead forecast variances for the models  $M^{(i)} , i > 1$ , are equal. For example, with prespecified  $\beta_1^{(1)} , \beta_2^{(1)} , \beta_1^{(2)}$  and  $V^{(1)}$ , i.e: given  $B^{(1)} , B^{(2)}$  and  $V^{(1)}$ ,  $R^{(1)}$  and  $\hat{Y}^{(2)}$  can be calculated. The outlier variance  $V^{(3)}$  can then be set so that  $V^{(3)} = \hat{Y}^{(2)} - f R^{(1)} f'$ . This gives  $\hat{Y}^{(3)} = \hat{Y}^{(2)}$ .



Since  $B^{(4)} = \text{diag}\{\beta_1^{(1)}, \beta^{(4)}\}$ ,  $\beta^{(4)}$  can be chosen so that  $\dot{Y}^{(4)} = \dot{Y}^{(3)}$ . In particular, defining  $r_i = f_i G_i C_{ii,t-1} G_i' f_i'$ ,  $i=1,2$  and  $r_{12} = f_1 G_1 C_{12,t-1} G_2' f_2'$ , we have

$$(\beta^{(4)})^{-1} r_2 + 2(\beta_1^{(1)} \beta^{(4)})^{-1} r_{12} + V^{(1)} = 2(\beta_1^{(1)} \beta_2^{(1)})^{-1} r_{12} + (\beta_2^{(1)})^{-1} r_2 + V^{(3)}.$$

This is a second degree equation and can be solved for  $(\beta^{(4)})^{-1}$ . This operates when the CUSUM first signals a change. If required during the single model phase, the variance  $V$  may be estimated on-line using either methods described in Chapter 5.

During the multiprocess phase the forecast distribution is often multimodal being the mixture of Normals. Point forecasts are then derived using a conjugate Normal loss function introduced by Lindley (1976). Further discussion of the use of such loss functions can be found in Smith, Harrison and Zeeman (1981) and Harrison and Smith (1980).

## 7.6. SUMMARY :

The principle of Management by Exception is discussed and a historical background of the forecasting systems is given in Section 2. The backward CUSUM statistic is reviewed in Section 3 and in Section 4, the NWB multiprocess models are introduced in the light of Harrison and Stevens multiprocess models. Finally all the above ideas are combined in Section 5 to give multiprocess models with CUSUMS and this is explained in some detail for a linear growth seasonal model.



## CHAPTER EIGHT

### APPLICATIONS

#### 8.1. INTRODUCTION :

This chapter is devoted to applications of the developed theory in a variety of situations. The NDBM's  $\{f, G, V, B\}$  are constructed using the principle of superposition since any multivariate Normal random vector can be decomposed into a linear combination of component multivariate Normal random vectors. This suggests that  $G = \text{diag}\{G_1, G_2, \dots, G_r\}$  where the block  $G_i$  is associated with a meaningful model component. Accordingly,  $f = [f_1, f_2, \dots, f_r]$  and  $B = \text{diag}\{\beta_1 I_1, \beta_2 I_2, \dots, \beta_r I_r\}$  each with proper dimensionality. This includes the case where  $G_i = \lambda_i$ , the  $\lambda_i$ 's being distinct eigenvalues of  $G$ . But for real observation processes, where complex eigenvalues are concerned, it is usual to consider conjugate pairs in the same block. That is, for a pair of conjugate complex eigenvalues  $(\lambda e^{i\omega}, \lambda e^{-i\omega})$  of multiplicity one, the adopted form is  $G = \lambda \begin{bmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{bmatrix}$ . This could represent a damped sine wave of period  $2\frac{\pi}{\omega}$  and would typically have a single associated discount factor  $0 < \beta < \lambda^2$ . The discount factors used throughout are not chosen according to any optimisation criterion. There might be room for further research here. Trigg and Leach(1967) have attempted to redefine Brown's discount factor as specific functions of sign and absolute one step ahead error forecasts. For an EWR  $\{P=1, G=1, \beta\}$ , with  $0 < \beta < 1$ , it is easily seen that in the limit, the adaptive coefficient  $a_t$  is  $1 - \beta$ . This gives  $m_t = \beta^t m_0 + (1-\beta) \sum_{i=0}^{t-1} \beta^i y_{t-i}$ , with  $(1-\beta)\beta^k$  as the weight corresponding to a data point that is  $k$  periods old. Comparing the average age of the data in an  $N$  period moving average,  $(1/N) \sum_{i=0}^{N-1} (N-1-i) = (N-1)/2$ , to the average age  $(1-\beta) \sum_{i=0}^{\infty} i \beta^i$ , based on the above EWR model, Montgomery and Johnson

(1976) have obtained the relationship  $\beta = (N - 1) / (N + 1)$ . Using this relation, Agnew (1982) suggested that  $0.33 \leq \beta \leq 0.78$ . Clearly, such low values of  $\beta$  give highly adaptive models with large lead time forecast variances, which would be totally unsuitable for such purposes as stock control and production planning. A rather more encouraging suggestion is that of Harrison and Johnston (1983) given by  $\beta = \frac{3N-1}{3N+1}$  where

$N$  represents age to half effect. That is,  $N$  is the time for the weight of a particular data point to halve in value. This leads to higher  $\beta$  values.

Apart from the discontinuity periods, the discount factors here are chosen more close to 1 such that the more stable the component the closer its discount factor is to 1. Experience shows model robustness against this choice.

In modelling discontinuity periods, the Modified NDBM's are used in order to protect model components information from unwanted interactions and the guard procedure on the observation variance described in 7.5 is employed.

For a straightforward application of a single NDBM to a data series which exhibits no major changes and outliers, see the US air passengers data set which is analysed in Chapter 2. Other selected series considered here, are :

- i) A simulated seasonal series with trend, level, seasonal changes, outliers and missing observations. This is to examine the performance in the phase of these major changes and discontinuities knowing the true underlying model. The data is analysed using both intervention and multiprocess NDBM's.
- ii) In order to test the performance of the multiprocess NDBM's vis the CUSUM multiprocessor, a medical data set concerning prescription charges is chosen. The data was previously analysed using Harrison-Stevens multiprocess models.
- iii) For a typical data set with an unknown and variable observation variance, the Road Death Series is chosen and the CUSUM multiprocessor is applied with an on-line estimation of the observation variance.

All the data sets are provided in the appendix.

## 8.2. SIMULATED SERIES

In order to examine model performance in phase of major changes and impulses with a minimum risk of misspecifications, artificial data is generated. The data is analysed using both an automatic method and using Intervention.

For an automatic way of dealing with these changes, a multiprocess model is used. Automation in analysing statistical data sets may not be a desirable property to aim for from a Bayesian point of view. However, multiprocess models have a wide range of applications in areas other than prediction of future outcomes and are especially valuable in the detection and classification of different types of changes in process components.

### 8.2.1. Simulation of the Data :

The artificial data is simulated by the superposition of three component series. These are an independent random noise  $v_t$ , a linear trend component  $\boldsymbol{\omega}'_{1,t} = [w_1, w_2]_t$  and a cyclic component  $\boldsymbol{\omega}'_{2,t} = [w_3, w_4]_t$ , represented by a single harmonic of period 12. The simulation is carried out using the model :

$$Y_t = f \boldsymbol{\theta}_t + v_t$$

$$\boldsymbol{\theta}_t = G \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

where  $f = [1, 0, 1, 0]$ ,  $G = \text{diag}\{G_1, G_2\}$ ,  $G_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ ,  $G_2 = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}$ . with  $\psi = \frac{\pi}{6}$ ,

$v_t \sim N[0; 400]$ ,  $\boldsymbol{\omega}'_t = [\boldsymbol{\omega}'_{1,t}, \boldsymbol{\omega}'_{2,t}] \sim N[0; \text{diag}\{13.175, 0.017, 1.14, 1.14\}]$  and

$\boldsymbol{\theta}'_0 = [75, 9, 85, 1]$ .

Accordingly, a series of 120 monthly observations is generated and the following major impulses and events are imposed:

- i) 200 is subtracted from the intermediate observation at  $t=32$ , in order to simulate an outlier;



- ii) immediately after  $t=36$  , the process ' deseasonalised ' level is reduced by 270 to give a jump ;
- iii) following  $t=60$ , the linear growth is reversed in sign from roughly 8 units per period to -8 , giving a slope change ;
- iv) following  $t=85$  , the linear growth is again reversed in sign and simultaneously the seasonal amplitude is increased by 50 % ;
- v) data points 113, 114 and 115 are eliminated to give a period of three missing observations.

### 8.2.2. INTERVENTION :

Intervention involves changing a routine or existing probability model often by introducing subjective information. In classical time series, interventions are specified through transfer functions ( Box and Tiao (1975)). In Bayesian Dynamic Models, intervention is achieved through transfer probability distributions which not only introduce an expected effect but also introduce an extra uncertainty associated with the change. The object of structuring a model, is to enable changes to be made to particular model components in such a way that leaves other components largely unaffected In the following example, a useful way in which additional uncertainty can be specified through the discount factors is illustrated. As explained earlier, the discount factors replace the role of the state random noise  $w_t$ .

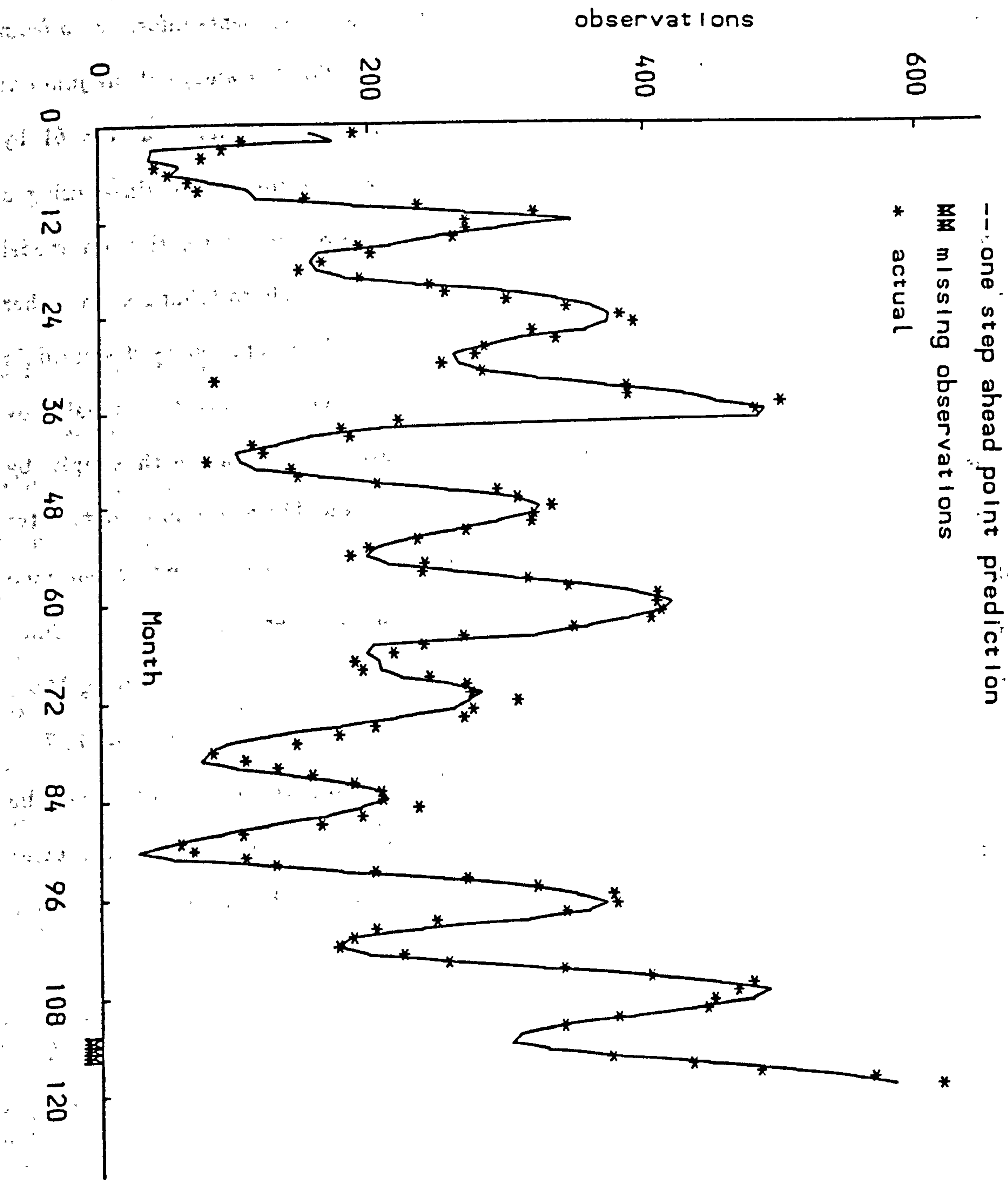
For the data simulated in 8.2.1, the NDBM  $\{f, G, V_t, B_t\}$  is applied with  $f, G$  and  $V$  as defined there and  $B = \text{diag}\{\beta_1, \beta_2, \beta_3, \beta_4\}_t$ . Apart from at times of intervention , the values  $\beta_1 = \beta_2 = 0.9$  and  $\beta_3 = \beta_4 = 0.95$  are used. No attempts to improve model performance are made by looking for 'optimal' discount factors. The values chosen, are rounded figures thought to be appropriate , bearing in mind that it is usually preferable to err on the side of underestimating discount factors, Harrison (1967). Initially , the same starting values for  $\theta_0$  are adopted but with a vague covariance matrix  $2000 I$ . When the major changes (i) to (v) are about to occur it is assumed that the type of forthcoming

event is known but that there is no available information on the size or even the sign of the coming change. Since, it is known that  $y_{32}$  is an outlier, the updating procedure treats it as a missing observation. This is to protect model components information from misspecifications that the outlier observation may provide. Foreknowledge of the jump at  $t=37$  is signaled by  $(\beta_1, \beta_2) = (0.9^{40}, 1)$  and for the trend growth change at  $t=61$  by  $(\beta_1, \beta_2) = (1, 0.9^{80})$ . In each instance, the updating distributions are obtained using a modified NDBM since, in practice, it is desirable to protect information on model components from the effect of imprecise descriptions of sharp changes in other components. At these intervention times a Modified NDBM is applied as described in 4.4.1. The simultaneous sudden change in trend and seasonality at  $t=86$  is signaled by  $B_{86} = \text{diag}\{1, 0.9^{80}, 0.95^{30} I\}$ . The three missing observations are dealt with simply by taking the posterior distributions of  $\theta_t$  for  $t=113, 114$ , and  $115$  as the prior parameter distribution  $(\theta_t | D_{112})$ . Fig.3 shows the observations and the corresponding one step ahead expectations in order to demonstrate the power of the intervention method. Since in the routine data generation  $V = 400$ , without major disturbances the limiting Mean Absolute Deviation (MAD) of the one step ahead forecast errors would be about  $18.7 = 0.8 \hat{Y}^4$ , where  $\hat{Y} = 400. / (\beta_1^2 \beta_2^2)$  (Theorem 6.4). The overall performance in terms of the MAD is found to be 20.76 after omitting the outlier  $e_{32}$ , the jump  $e_{37}$  and the three missing errors  $e_{113}$ ,  $e_{114}$  and  $e_{115}$ . The performance in terms of the MAD for each year is given below.

YEAR	1	2	3	4	5	6	7	8	9	10
MAD	39.5	20.3	20.5	20.3	15.0	17.2	17.4	20.1	21.5	15.8



FIG. 3. AN EXAMPLE OF INTERVENTION ON AN ARTIFICIAL DATA SET



### 8.2.3. Multiprocess Models - The Artificial Data :

For an automatic way of dealing with the major changes in the series , it is assumed that the series is monthly with an additive linear growth and one harmonic seasonal component. Given this structure, the possible changes in the series which could be considered are trend change, outlier, seasonal changes and/or combinations of them (  $2^3$  possibilities ) alongside of the mother model. However, successive operations of the main changes may reasonably model the combined changes, if any. Clearly, the computer storage and running time increases exponentially with the number of models considered. This suggests that a fewer number of alternative models should be considered.

For a rough comparison with intervention results. an NDB multiprocess model is constructed with four models comprising: A basic or mother model with trend discount factor  $\beta_1=0.9$  and seasonal discount factor  $\beta_2=0.95$  and observation variance  $V = 400$ , all as specified in the basic intervention model, a trend change model with discount factors  $\beta'_1=0.02$  and  $\beta'_2=\beta_2$ , an outlier and a seasonal change model. The seasonal change discount factor and the variance for the outlier model are found from the observation variance controlling rule for the alternative models defined in 7.5. Transition probabilities from models at time  $t$  to each of the mother, trend, outlier and seasonal change models at time  $t+1$  are taken as 0.8, 0.095, 0.1 and 0.005 respectively.

Given the same initial settings as in the intervention case, the data is then analysed using NDB multiprocess models without and with the CUSUM controller. In the latter case,  $L_0$  and  $\alpha$  are taken as 2.0 and 0.5 respectively with a threshold probability of 0.8 for switching back to single model operation.

As is to be expected, unlike the intervention model, multiprocess models need more time to recognise changes and make proper model adjustments. The high forecasting errors observed for the years 3, 4, 6 and 8 , may partly be due to that and partly due to the selection of the alternative models. Level and growth changes are combined in the trend change model, while, seasonal and growth change are modelled together. This is to

keep the process of alternative model selection vague and simple. A summary of the models performance without a CUSUM is presented in Fig.4 and Fig.5 presents that with the CUSUM statistic. The overall MAD's after removing the errors  $e_{32}, e_{37}, e_{38}, e_{39}, e_{113}, e_{114}$  and  $e_{115}$  were found to be 25.03 and 22.74 respectively. The performance in terms of MAD for each year in both cases is given in the following table for comparison.

**Multiprocess Models - Simulated Data ( with and without CUSUMS)**

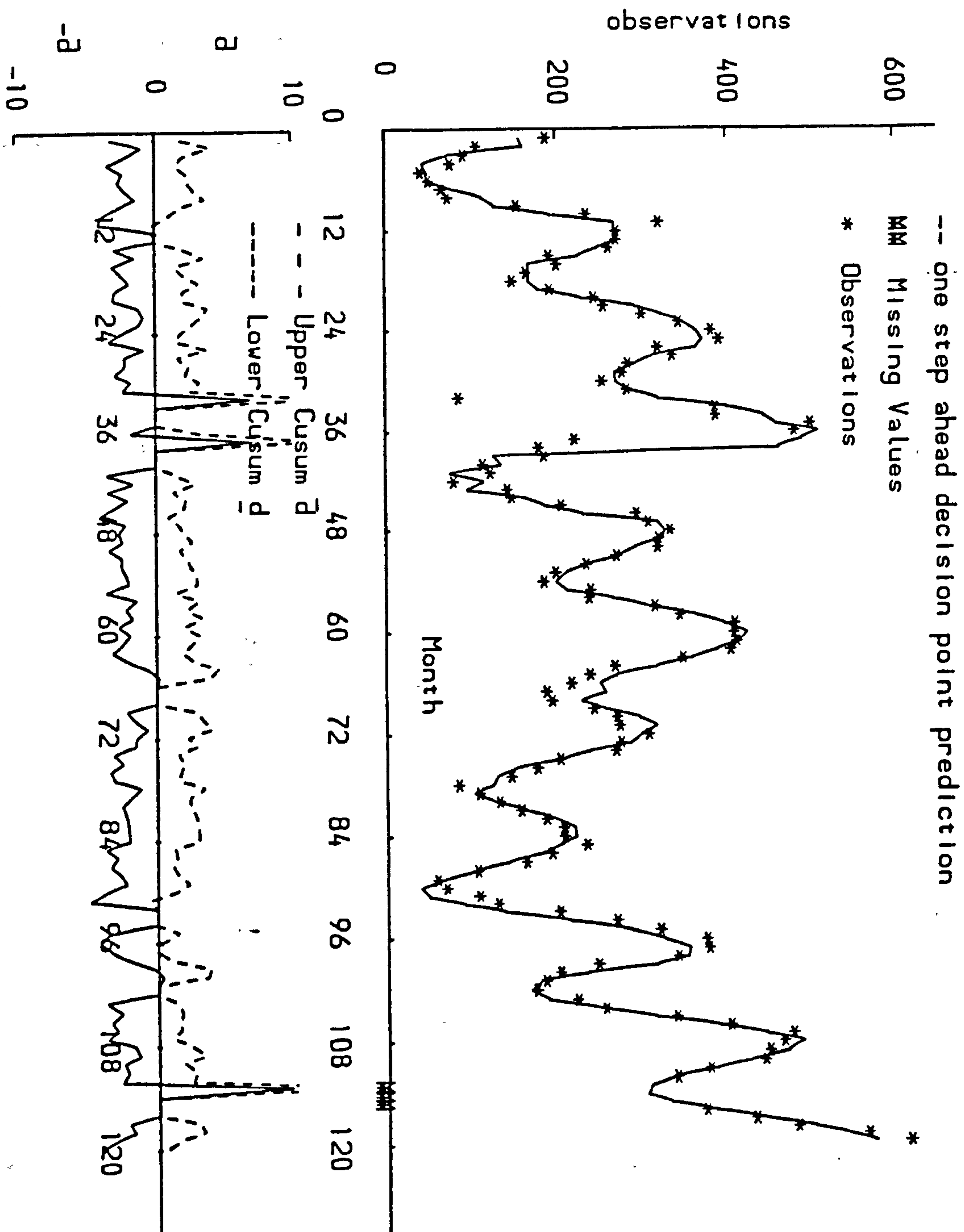
YEAR	1	2	3	4	5	6	7	8	9	10
without CUSUM's	32.8	21.3	34.7	40.4	15.6	22.5	16.5	26.3	20.9	19.3
with CUSUM's	26.9	17.9	23.9	29.1	13.9	28.0	14.9	32.9	22.4	17.5

It is easily observed from the above table, that the multiprocess model with CUSUM's is to be preferred in both terms of performance and computer storage and running time. Moreover, it is interesting to note that, the order of preference among the models, is in accordance with the amount of information available to be used for the model construction. The intervention model would be the best choice when all the information about the structural changes and their occurrence times are known. The second best choice would be, the multiprocess models with CUSUM's when a preferred model is known. Finally, if no information about a preferred model and the types of change and their occurrence times is known, the multiprocess model without the CUSUM statistic would be the candidate.

[illegible]



FIG. 5. SIMULATED SERIES -CUSUM



### 8.3. THE PRESCRIPTION SERIES :

#### 8.3.1. The Data :

This is a monthly medical data set giving the number of prescriptions for five years starting from March 1966. the figures taken are normalised according to the number of effective working days in the month. This is to compare the result with the analysis of Harrison and Stevens who previously used multiprocess models. The data is strongly seasonal for which a constant observation variance is reasonably assumed.

It is observed that increased prescription charges in June 1968 caused a major change in the level and an influenza epidemic in December 1970 'caused' an outlier. However, for the purpose of demonstrating multiprocess modelling, it is assumed that these events are not known. Consequently they are not anticipated but, are dealt with automatically together with other unobserved changes. The data is analysed using Modified NDB multiprocess models without and with the CUSUM statistic using the same initial prior information given as

$$(\theta_0 | D_0) = \left( \begin{bmatrix} \mu_1 \\ \mu_2 \\ s \end{bmatrix} | D_0 \right) \sim N \left( \begin{bmatrix} 22 \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} 10 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 25I_{11} \end{bmatrix} \right)$$

signifying a weak prior with no growth and no seasonal pattern. In both cases only two types of major disturbances are considered, namely, sharp trend changes and outliers.

#### 8.3.2. NDBM Multiprocess Models - Known Observation Variance :

Here the routine model has a linear growth and full seasonal components with corresponding discount factors  $\beta_1 = 0.95$  and  $\beta_2 = 0.975$ . The discount factors for the trend change model are  $(\beta_1, \beta_2) = (0.02, 0.975)$  and the model transition probabilities are constant throughout so that:

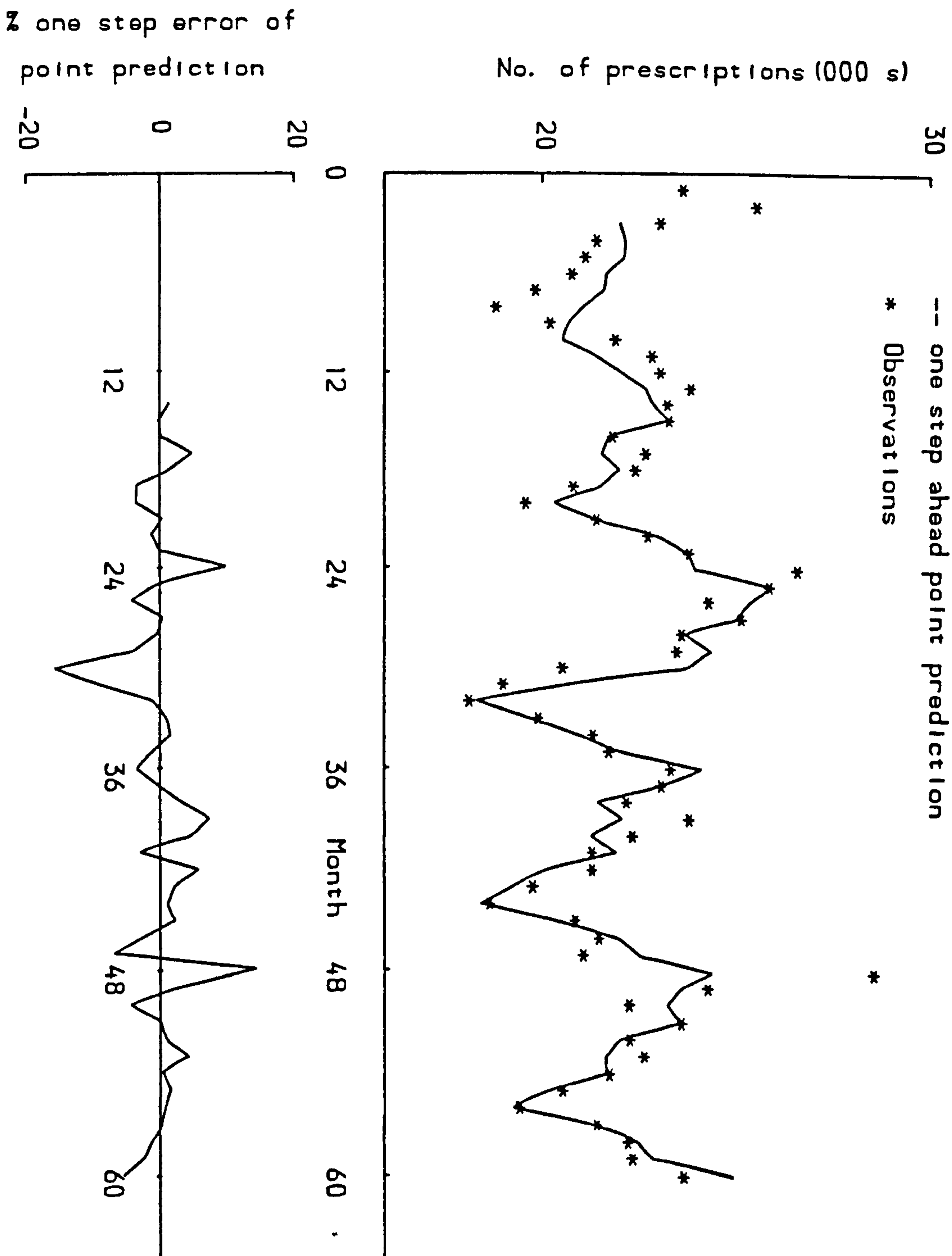
$$\pi_2 = P\{M_{t+1}^{(2)} | M_t^{(i)}\} = 0.05 ,$$

$$\pi_3 = \{M_{t+1}^{(3)} | M_t^{(i)}\} = 0.025 ; i = 1, 2, 3$$

where the  $M^{(i)}$  's are: Mother , Outlier and Trend change models respectively. The observation variance is estimated as 0.36. the model operation is as given in 7.4.

Given the posterior model information at time  $t-1$  as  $\{(\mathbf{m}_i; \mathbf{C}_i); P_i\}_{t-1}$  ( $i = 1, 2, 3$ ), the probability that at time  $t-1$  model  $i$  operated and that at time  $t$  model  $j$  operates is  $\pi_j P_j$  and the corresponding prior parametric distribution is  $(\boldsymbol{\theta} | M_i, M_j, D_{t-1}) \sim N[\hat{\boldsymbol{\theta}}_{ij}; \mathbf{R}_{ij}]$  where the  $\mathbf{R}_{ij}$  's are calculated according to the Modified NDBM rules. In order to control the response of the models to the outliers and jumps, the outlier variance is chosen so that  $\hat{Y}_2 = \hat{Y}_3$ . Point predictions are obtained using the conjugate Normal Loss function introduced by Lindley (1976) and plotted along with the observations in Fig.6 with the percentage one step ahead error forecasts. The first 12 observations are used for trend and seasonal pattern recognition. the model has recognised a minor unobserved shift at month 24 . the increase of prescription charges at month 30 has caused a negative error of -15% and is followed by an error of -5% as the uncertainty between outlier and sharp trend changes is resolved. the influenza epidemic at month 48 is properly identified as an outlier. The performance in terms of MAD, for the last four years, is tabulated in Section 8.3.3. This is for a direct comparison with the results obtained using multiprocess models with CUSUM's.

FIG. 6. THE PRESCRIPTION SERIES, MULTIPROCESS MODELLING



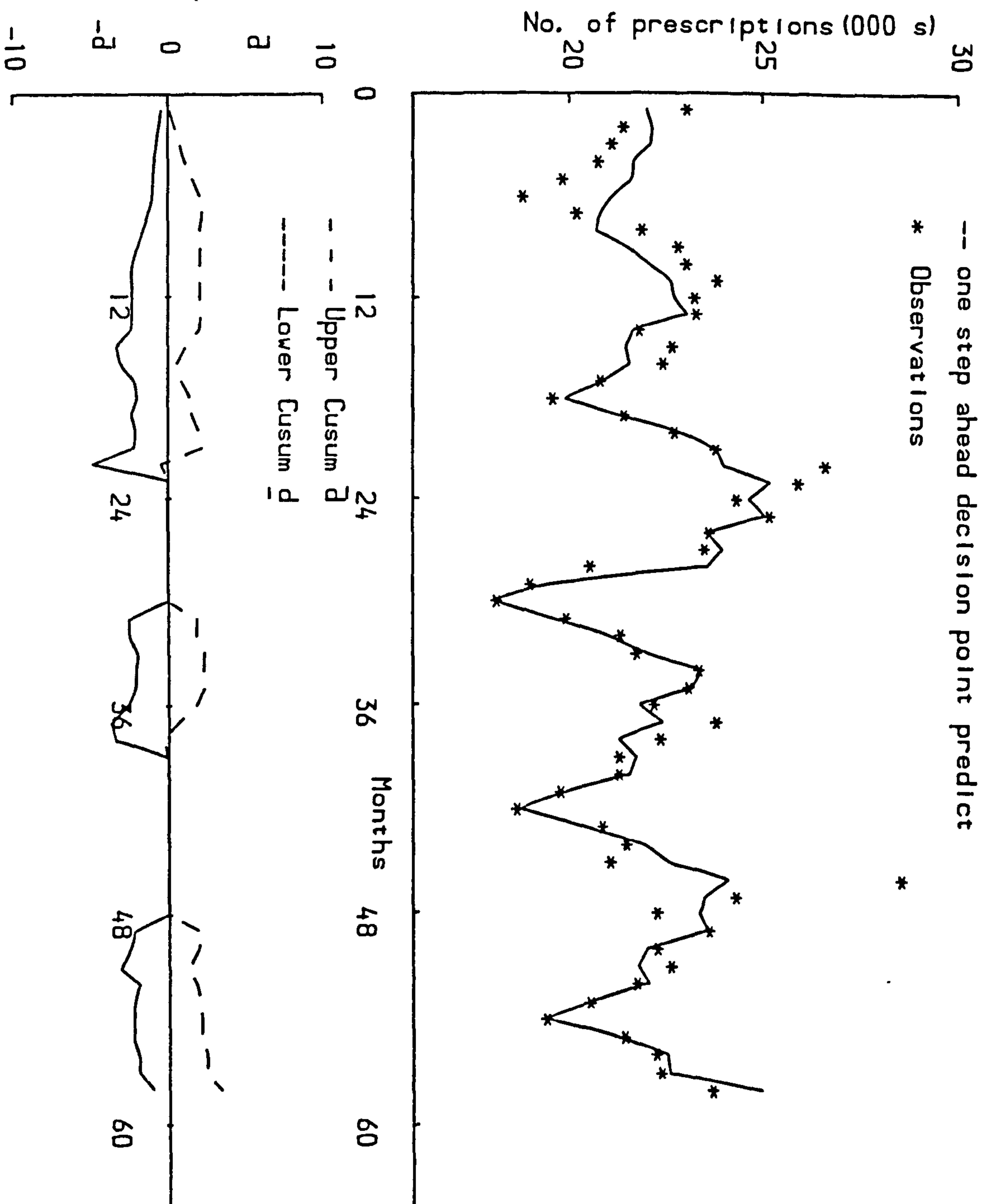


### 8.3.3. The CUSUM Multiprocessor - Known Observation Variance:

The same model specifications described in 8.3.2 are resumed here with the Backward CUSUM statistic with initial values  $L_0=2.0$  and  $\alpha=0.5$ . Predictions are based on the Mother model performance until the CUSUM monitor signals a change, at which time all the three models start operating interactively until a threshold probability of 0.98 is regained for the Mother model. during the Mother model performance other models run in parallel as described in 7.5 as preparatory arrangements for coming changes. The model performance is summarised in Fig.7 together with the upper and lower Backward CUSUM monitors. It can be seen that all the changes are properly identified and that the performance is slightly better than that of 8.3.2 while the process time is reduced nearly by 2/3. In order to compare the performance with that of the multiprocess models without the CUSUM statistic, the MAD for the last four years is tabulated below for the two models.

YEAR	1	2	3	4
without CUSUM's	0.63	0.52	1.79	0.47
with CUSUM's	0.57	0.46	0.7	0.34

FIG. 7. THE PRESCRIPTION SERIES



## 8.4. ROAD DEATH SERIES :

### 8.4.1. The Data :

This is a series of 38 observations representing quarterly road deaths in U.K. for the years 1960-1969. It can be seen from Fig.8 that the main observed discontinuities present in the series are the outlying observation in the first quarter of 1963 due to a cold icy winter preventing traffic using many roads and the trend change in 1967 due to the introduction of breathalyser. Generally , a high variation of the observation error variable can be observed. this suggests that an on-line estimation of the observation variance is more appropriate than a fixed and global estimate. Using longer data sets, a relation between the number of road deaths and industrial activity is evident. The death rate rises during a boom period when more traffic is on the roads and falls during slump periods. This effect could be accommodated in a model and would lead to more reliable predictions. However, the analysis here is for demonstration purposes and no attempt is made to relate road deaths and industrial productions.

### 8.4.2. the NDB Multiprocess Models with CUSUM 's :

In this analysis, Modified NDB multiprocess models are used with CUSUM's. The alternative models assumed are: trend change, outlier and seasonal changes. The main model sets the discount factors  $\beta_1=0.8$  and  $\beta_2 = 0.9$ . These figures are lower than the ones used in the precision example and reflect the quarterly data. The trend change discount factors are  $(\beta_1', \beta_2) = (0.8^4, \beta_2)$ . As before, the discount factors for the seasonal change model and the observation variance for the outlier model are found using the control rule for the alternative models observation variance defined in 7.5. The variance of the main model is estimated on-line assuming it to be proportional to the expected number of deaths, i.e.

$$V_t = \alpha E\{Y_t | D_{t-1}\}.$$

$\alpha$  is estimated on-line using the power law defined in 5.4 with initial values  $X_0 = 10$ . and

$\eta_0 = 20$ . The model transition probabilities are 0.7899, 0.1, 0.11 and 0.0001 with the same order as in 8.2.3. The CUSUM initial parameters where  $L_0 = 2.0$  and  $\alpha = 0.5$  and the threshold return probability is 0.8.

For this model:

$$f = [1, 0, 1, 0, 1] \text{ and } G = \text{diag}\{G_1, G_2\}$$

$$\text{where } G_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } G_2 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

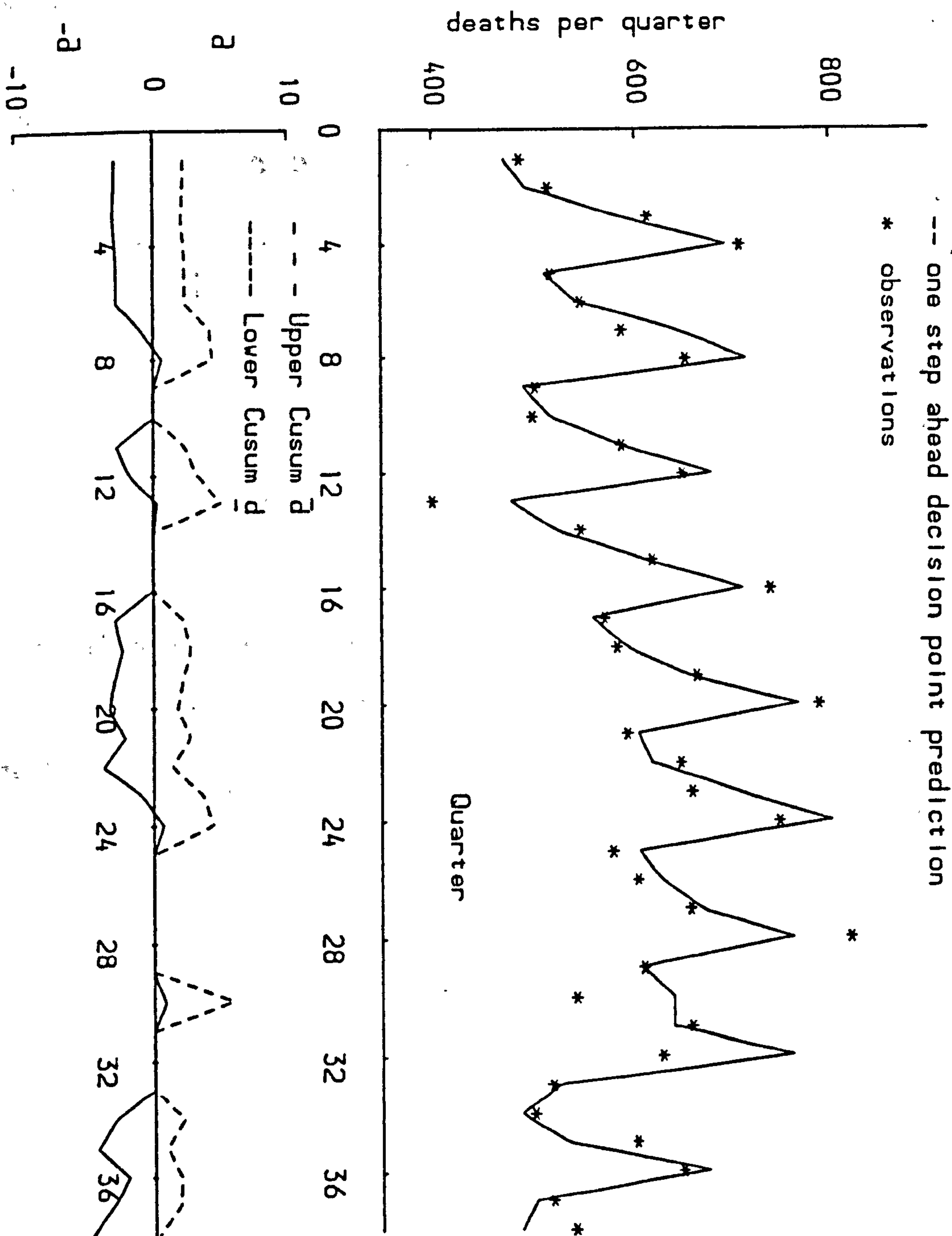
A weak prior distribution was set initially as

$$\left( \begin{bmatrix} \mu_1 \\ \mu_2 \\ s \end{bmatrix} | D_0 \right) \sim N \left[ \begin{bmatrix} 550 \\ 0 \\ 90 \\ -50 \\ 30 \end{bmatrix}; \text{diag}\{1000, 100, 1000 I_3\} \right]$$

Finally, a summary of the results and the performance is presented in Fig.8 showing that all the prescribed changes are dealt with successfully. The summary on the CUSUM statistic shows periods of different model operation and the direction of changes at the model breakdown points.



FIG. 8. ROAD DEATH SERIES



### 8.5. SUMMARY :

In this chapter a number of applications of the Normal Bayesian Models based on the principle of parsimony through discounting and Management by Exception are presented. The first application is made by examining an artificially generated data using intervention at the points of discontinuity. This is presented in 8.2.2. In 8.2.3 the same data is analysed using both multiprocess models with and without CUSUM's. In Section 3, the models are applied to a real data concerning prescription charges where a constant observation variance is assumed. For an on-line estimation of the observation variance, quarterly data is chosen in Section 4 concerning the number of road deaths in U.K.. Appropriate figures are presented in each case to summarise the model performance.

## CHAPTER NINE

### DISCUSSION AND FURTHER RESEARCH

The Bayesian framework in statistics, is most promising and logical among existing statistical techniques. In modelling and predicting future outcomes, it requires supervision and interaction of the modellers to accommodate, on-line, any environmental or external effects that are not anticipated.

The pioneering work of Harrison and Stevens(1976) has provided applied statisticians with a base for analysing series arriving sequentially with time, away from static and limiting predictor models. The DLM's have seen a number of successful applications by Harrison and Stevens(1975), Smith (1983) and Smith and West(1983). However, as pointed out in Chapter 3 of this study, both the observation and state covariance matrices are ambiguous, not scale invariant and not parsimonious in the sense of Roberts and Harrison (1984). These problems have caused practitioners considerable difficulties in the estimation problem and diverted them to the use of other less constructive models.

The principle aim of this study is to replace the state error variance by a small number of discount factors. This gives models which enjoy the principle of parsimony within the Bayesian framework. Discount factors can more easily be set, they are invariant under linear transformations of scale, not ambiguous, and models based on the discount principle are generally parsimonious and robust. The discount principle is also aimed at demonstrating and publicising the potential of Bayesian modelling in practical applications, in particular, the study of processes with certain types of discontinuity. After providing the basic principle of discounting within the classical point estimation framework and testing its efficiency through comparison with DOUBTS and ARIMA models applications, the drawbacks and limitations of the Bayesian DLM's are pointed out. The discount principle is then carried out to construct NWBM's replacing the



DLM's. This class of models has many interesting subclasses and many existing and well known classical models are retained in the sense that they have the same limiting forecast functions as special constant NDBM's. However the Bayesian facilities are more extensive. Two methods are given for on-line estimation of the observation variance. This is essential especially for processes with high stochastic variations and those that exhibit sudden changes and outliers. In these cases, multiprocess models are advisable. The controlling rule for the observation variances is advisable since the variance governs the model likelihood ratios. The use of CUSUM statistic provides an overall improvement in the efficiency, computer storage and running time problems. The NDBM's allow a simple and easy way of communication and intervention in phases of major disturbances. Almost all types of major disturbances that are common in time series processes are present in the artificially generated data set. Even less disturbed series of that kind are often avoided by statisticians, see Chatfield (1978). The discount principle has simplified intervention with different components, as the example in 8.2.2 demonstrates. All types of change are detected successfully. However, in the analysis of real data sets advance information on major disturbances is often missing, in which case it is useful to adopt multiprocess models. Clearly, as is to be expected, the resulting analysis from the multiprocess models shown in Fig.4 is less successful than those from the intervention models. Apart from the missing observations, no information on the disturbances is fed into the multiprocess models. More efficient results are obtained using the multiprocess models with CUSUM's where a little more information is provided by assuming the existence of a particular model representing the process. The models are also applied to real data sets. These are used to demonstrate the efficiency of the models in dealing with certain types of discontinuity occurring when the data is fairly stable and also when high observation noise is present. This causes a delay in the task of recognising changes. The results from all applications dealt with are promising. In particular, when multiprocess models are called for, the CUSUM statistic is recommended for efficiency and economy and the Modified NDBM's protect component influences and misspecifications when major



disturbances are present. More applications can be found in Ameen and Harrison (1983 a , b , c ). In all cases the underlying model parameters have been given physical meanings and simple transformations are provided to transfer information from or to other practical applications of interest. It can be argued that the amount of further developments and exploitation in these models is proportional to the amount of effort spent in developing the existing and less profound models. The following lists a number of suggestions for further research:

- i- The models deal with processes defined only on the entire real line with the Normality assumptions so that successive estimates are obtained using Kalman Filters recurrence relations. However, in many real life problems , processes are well defined on bounded sample spaces and do not cover the real line sensibly, in which case, these models may provide estimates outside their feasible region. These points seems to be the most promising and demands exploitation.

Smith (1979) and Souza and Harrison (1979) have extended the DLM's to include non Normal Steady State models. These ideas are combined with the discount principle, Ameen (1983 b), to provide Generalised Bayesian Entropy models.

- ii- The forecast functions are specified using the design and transition matrices.

It is important to develop methods that provide more automation in model identification and a proper Bayesian on-line parameter learning procedure will improve the performance. Some considerable success has been achieved by Migon and Harrison (1983) considering non linearity and non Normality of the processes.

- iii- The choice of discount factors is left to the modellers and work needs to be done in developing methods for on-line estimation. The generalised EWR and the limiting ARIMA models obtained in chapter 6 have restricted parameters as

pointed out by Godolphin and Stone (1980) for the DLM's in which they suggest the use of singular transition matrices. Also, with lower discount factors, the uncertainty of lead time forecast distributions explode rapidly providing less reliable long term predictions.

- iv- Generalising the models to include more correlation structures will provide a wider range of applications.
- v- The limiting results obtained are mostly based on specified canonical representations and more general results are possible.
- vi- In a general context, more applications of the theory in different fields of interest are needed especially when the process is subject to a dynamic development as almost always is the case. The NDBM's replace the popular classical regression models and provide an overall improvement in the analysis. Some applications on this topic are given by Harrison and Johnston(1983).

## APPENDIX

### U.S.AIR PASSENGERS DATA

YEAR	951	952	953	954	955	956	957	958	959	960
JAN	145	171	196	204	242	284	315	340	360	417
FEB	150	180	196	188	233	277	301	318	342	391
MAR	178	193	236	235	267	317	356	362	406	419
APR	163	181	235	227	269	313	348	348	396	461
MAY	172	183	229	234	270	318	355	363	420	472
JUN	178	218	243	264	315	374	422	435	472	535
JUL	199	230	264	302	364	413	465	491	548	622
AUG	199	242	272	293	347	405	467	505	559	606
SEP	184	209	237	259	312	355	404	404	463	508
OCT	162	191	211	229	274	306	347	359	407	461
NOV	146	172	180	203	237	271	305	310	362	390
DEC	166	194	201	229	278	306	336	337	405	432

### SIMULATED DATA

YEAR	1	2	3	4	5	6	7	8	9	10
JAN	189	270	392	221	318	411	273	233	377	449
FEB	108	261	318	179	317	404	267	192	340	444
MAR	93	192	335	185	269	347	202	163	246	377
APR	77	201	283	114	234	267	176	106	202	338
MAY	42	166	276	122	198	238	146	59	185	
JUN	52	150	253	80	185	216	84	69	175	
JUL	67	193	282	143	239	187	108	108	222	
AUG	75	244	86	148	237	193	132	130	254	373
SEP	155	255	387	205	314	242	157	201	338	433
OCT	236	300	388	292	343	269	187	268	402	483
NOV	320	343	501	307	409	272	206	319	478	567
DEC	270	382	482	331	408	306	207	375	467	617



PRESCRIPTION DATA

YEAR	1966	1967	1968	1969	1970
JAN		23.9	25.9	23.1	24.3
FEB		23.3	24.4	22.2	22.3
MAR	23.1	23.3	25.2	23.8	23.6
APR	21.4	21.8	23.6	22.4	22.3
MAY	21.1	22.7	23.5	21.3	22.6
JUN	20.8	22.4	20.5	21.3	21.7
JUL	19.8	20.8	19.0	19.8	20.5
AUG	18.8	19.6	18.1	18.7	19.4
SEP	20.2	21.4	19.9	20.8	21.4
OCT	21.9	22.7	21.3	21.5	22.3
NOV	22.8	23.8	21.7	21.0	22.4
DEC	23.1	26.6	23.4	28.6	23.7

ROAD DEATH DATA

QUAR	960	961	962	963	964	965	966	967	968	969
1	486	516	501	400	570	592	578	610	518	518
2	514	546	499	547	582	648	604	542	499	541
3	614	587	587	619	664	660	658	659	603	
4	710	653	650	742	790	751	822	629	650	



## REFERENCES

- [1] AGNEW, R. A. (1982). Econometric forecasting via discounted least squares. Naval Research Logistics Quart., Vol. 29, No. 2, 291-302.
- [2] AMEEN, J. R. M. (1983 a). Contribution to discussion of the paper by E. T. Jaynes. Second International Meeting on Bayesian Statistics, Valencia, Spain, Sept. 1983.
- [3] AMEEN, J. R. M. (1983 b). Generalised Bayesian entropy models and forecasting. Warwick Res. Rep. 37.
- [4] AMEEN, J. R. M. (1983 c). Contribution to discussion of the paper by M. West. Second International Meeting on Bayesian Statistics, Valencia, Spain, Sept. 1983.
- [5] AMEEN, J. R. M. and HARRISON, P. J. (1983 a). Discount weighted estimation. J. of Forecasting. ( to appear).
- [6] AMEEN, J. R. M. and HARRISON, P. J. (1983 b). Normal discount Bayesian models. Invited paper for the second International Meeting on Bayesian Statistics, Valencia, Spain, Sept. 1983.
- [7] AMEEN, J. R. M. and HARRISON, P. J. (1983 c). Discount Bayesian multiprocess modelling with CUSUM's. Proceedings of International Time Series Conference, Nottingham, (O.D. Anderson ed.), 1983. North Holland.
- [8] ANDERSON, O. D. (1977). A commentary on " A survey of Time Series ". International Statist. Rev., 45, 273-297.
- [9] ASTROM, K. J. (1970). Introduction to Stochastic Control Theory. Academic Press, Inc., New York.
- [10] BARNARD, G. A. (1959). Control charts and stochastic processes. J. R. Statist. Soc. B, 21 , 239-270.
- [11] BISSELL, A. F. (1969). Cusum techniques for quality control ( with discussion ) . J.

- R. Statist. Soc. C, 18, 1-30.
- [12] BOX, G. E. P. and JENKINS, G. M. (1970). Time Series Analysis, Forecasting and Control: San Fransisco, Holden Day.
  - [13] BOX, G. E. P. and TIAO, G. C. (1975). Intervention analysis with applications to economic and environmental problems. J. Amer. Statist. Ass., 70, 70-79.
  - [14] BROWN, R. G. (1963). Smoothing, Forecasting and Control. San Fransisco: Holden Day.
  - [15] BROWN, R. G. (1983). The balance of effort in forecasting. J. of Forecasting. Vol.1, No.1, 49-53.
  - [16] CANTARELIS, N. and JOHNSTON, F. R. (1983). On-Line variance estimation for the steady state Bayesian forecasting model. J. Time Series An., Vol. 3, No. 4, 225-234.
  - [17] CHATFIELD, C. (1978). The Holt-Winters forecasting procedure. App. Statist., 27, No.3, 264- 279.
  - [18] De GROOT, M. H. (1970). Optimal Statistical Decisions. New York, Mc Graw-Hill.
  - [19] DOBBIE, J. M. (1963). Forecasting predict trends by exponential smoothing. Opns. Res. 11, 908-918.
  - [20] EWAN, W. D. (1963). When and how to use Cusum charts. Technometrics, 5, 1-22.
  - [21] EWAN, W. D. and KEMP, K. W. (1960). Sampling inspection of continuous processes with no autocorrelation between successive results. Biometrika, 47, 363-371.
  - [22] GATHERCOLE, R. B. and SMITH, J. Q. (1983). A dynamic forecasting model for a general class of discontinuous time series. Proceedings of International Time Series Conference, Nottingham,( O.D. Anderson, ed.). North Holland.
  - [23] GEBEL, A. (ed.).(1974). Applied Optimal Estimation. MIT Press, Cambridge.
  - [24] GODOLPHIN, E. J. and HARRISON, P. J. (1975). Equivalent theorems for

- polynomial projecting predictors. J. R. Statist. Soc. B, 37, 205-215.
- [25] GODOLPHIN, E. J. and STONE, J. M. (1980). On the structural representation for polynomial projecting models based on the Kalman filter. J. R. Statist. Soc. B, 42, 35-46.
- [26] HARRISON, P. J. (1965). Short-term sales forecasting. J. R. Statist. Soc. C.(Appl. Statist. ), 15, 102-139.
- [27] HARRISON, P. J. (1967). Exponential smoothing and short-term sales forecasting. Man. Sci., 13, 821-842.
- [28] HARRISON, P. J. and AKRAM, M. (1983). Generalised exponentially weighted regression and parsimonious dynamic linear models. International Conference held at Valencia, ( O. D. Anderson ed.), North Holland.
- [29] HARRISON, P. J. and DAVIES, O. L. (1964). The use of Cumulative sum ( Cusum ) techniques for the control of routine forecasts of produce demand. Oper. Res.(J. O. R. S. A. ) 12, 325-333.
- [30] HARRISON, P. J. and JOHNSTON, F. R. (1983). A regression method with non stationary parameters. Warwick Res. Rep. 35. ( Submitted to J. of O. R. ).
- [31] HARRISON, P. J., LEONARD, T. and GAZARD, T. N. (1977). An application of multivariate hierarchical forecasting. Paper to R. Statist. Soc. Ind. Appl. and Res. Section Conference, Manchester. Also Warwick Res. Rep. 15.
- [32] HARRISON, P. J. and PEARCE, S. F. (1972). The use of trend curves as an aid to market forecasting. Ind. Mark. Manage., 2, 149-170.
- [33] HARRISON, P. J. and SCOTT, F. A. (1965). A development system for use in short term sales forecasting investigations. Paper to Ann. Conf. O. R. Soc. and Special O. R. Soc. Meeting. Also Warwick Res. Rep. No. 26.
- [34] HARRISON, P. J. and SMITH, J. Q. (1980). Discontinuity, decision and conflict. Proc. of the first International Meeting on Bayesian Statistics, Valencia, Spain ( ed. Birnardo et al.) May 1979.



- [35] HARRISON, P. J. and STEVENS, C. F. (1971). A Bayesian approach to short-term forecasting. *Oper. Res. Quart.*, 22, 341-362.
- [36] HARRISON, P. J. and STEVENS, C. F. (1975). Bayesian forecasting in action : Case studies. *Warwick Res. Rep.* 14.
- [37] HARRISON, P. J. and STEVENS, C. F. (1976). Bayesian forecasting ( with discussion ). *J. R. Statist. Soc. B.* 38. 205-247.
- [38] HENDERSON, C. R. . KEMPTHORNE, O., SEARL, S. R. and KROSIGK, C. M. (1959). The estimation of environmental and genetic trends subject to culling. *Biometrika*, 15, 192- 218.
- [39] HOLT, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. *Carnegie Inst. Technol. Res. Memo.* No. 52.(NONR 760(01)).
- [40] JAYNES, E. T. (1983). Highly informative priors: the effect of multiplicity of inference. Invited paper for the Second International Meeting on Bayesian Statistics, Valencia, Spain, Sept.1983.
- [41] KALMAN, R. E. (1963). New methods in Wiener filtering theory. In *Proceedings of the first Symposium on Engineering Application of Random Function Theory and Probability.* ( J. L. BOGDANOFF and F. KOZIN. eds.) New York : Wiely.
- [42] KALMAN, R. C. and BUCY, R. S. (1961). New results in linear filtering theory. *J. Basic Eng.*, 83, 95-108.
- [43] KENDALL, M. , STUART, A. and ORD, J. K. (1983). *The Advanced Theory of Statistics. Vol.3, 4-th ed.*, Charles Griffin & Company Limited.
- [44] LEE, T. T. (1980). A direct approach to identify the noise covariances of the Kalman filtering. *IEEE. Trans. Automatic Control*, Vol. 1, AC-25,841-842.
- [45] LINDLEY, D. V. (1976). A class of utility functions. *Ann. Statist.* 4,1-10.
- [46] LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for linear models. *J. R. Statist. Soc. B*, 34 ,1-41.



- [47] MAKOV, UDI E. (1983). Approximate Bayesian procedures for dynamic linear models in the presence of jumps. *The Statistician* 32, 207-213.
- [48] MAYBECK, P. S. (1982). *Stochastic Models, Estimation and Control*. Vol.2, Academic Press, New York.
- [49] Mc KENZIE, E. (1976). An analysis of general exponential smoothing. *Oper. Res.* 24, 131-140.
- [50] MIGON, H. S. and HARRISON, P. J. (1983). An application of non-linear Bayesian forecasting to television advertising. Contributed paper to the Second International Meeting on Bayesian Statistics, Valencia, Spain. Sept.1983.
- [51] MONTGOMERY, D. C. and JOHNSON, L. A. (1976). *Forecasting and Time Series Analysis*, McGraw-Hill, New York.
- [52] MUTH, E. J. (1981). Forecasting with polynomials by a generalised recursive updating method. Paper to First Inter. Symp. on Forecasting. Quebec City, Canada.
- [53] O'HAGAN, A. (1979). On outlier rejection phenomenon in Bayes inference. *J. R. Statist. Soc. B*, 41,358-367.
- [54] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika*,41 ,100-115.
- [55] PRIESTLEY, M. B. (1980). State-dependent models: A general approach to non-linear time series analysis. *Time Series Analysis*, 1 ,47-71.
- [56] ROBERTS, S. A. and HARRISON, P. J. (1984). Parsimonious modelling and forecasting of seasonal time series. *Eur. J. Oper. Res.*, 16, 365-377.
- [57] SMITH, A. F. M. and COOK, D. G. (1980). Straight lines with a change point : A Bayesian analysis of some renal transplant data . *Applied Statist.* 29, 180-189.
- [58] SMITH, A. F. M. and WEST, M. (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics*, 39 ,No.4 , 867-878.
- [59] SMITH, J. Q. (1977). *Problems in Bayesian Statistics Relating to Discontinuous Phenomenon, Catastrophe Theory and Forecasting*. Ph.D. Thesis, Univ. of Warwick.

- [60] SMITH, J. Q. (1979). A generalisation of Bayesian steady forecasting model. J. R. Statist. soc. B, 41 , 378-387.
- [61] SMITH, J. Q. (1983). Forecasting accident claims for an assurance company. The Statistician, 32, 109-115.
- [62] SMITH, J. Q. , HARRISON, P. J. and ZEEMAN, E. C. (1981). The analysis of some discontinuous decision processes. Eur. J. Oper. Res. 7, 30-43.
- [63] SOUZA, R. C. (1978). A Bayesian Entropy Approach to Forecasting. Ph. D. Thesis, Univ. of Warwick.
- [64] SOUZA, R. C. (1981). A Bayesian entropy approach to forecasting : The multistate model. In Time Series Analysis ( Houston Tex. ) North Holland , 535-542.
- [65] SOUZA, R. C. and HARRISON, P. J. (1979). Steady state system forecasting : A Bayesian entropy approach. Warwick Res. Rep. 33.
- [66] STEVENS, C. F. (1974). On the variability of demand for families of items. Oper. Res. Quart., 25, 411-420.
- [67] TRIGG, D. W. and LEACH, G. A. (1967). Exponential smoothing with adaptive response rate. Oper. Res. Quart., 18, 53-64.
- [68] VAN DOBBEN DE BRUYN, C. S. (1968). Cumulative Sum Tests: Theory and Practice. London : Griffin.
- [69] WALD, A. (1947). Sequential Analysis. John Wiley & Sons. New York.
- [70] WEST, M. (1982). Aspects of recursive Bayesian Estimation. Ph. D. Thesis , Univ. of Nottingham.
- [71] WHITTLE, P. (1965). Recursive relations for predictors of non-stationary processes. J. R. Statist. Soc. B, 27, 523-532.
- [72] WHITTLE, P. (1969). A view of stochastic control theory. J. R. Statist. Soc. A, 132, 320-334.
- [73] WINTERS, P. R. (1960). Forecasting sales by exponentially weighted moving

- averages. Man. Sci., 6 , 324.
- [74] WOODWARD, R. H. and GOLDSMITH, P. L. (1964). Cumulative Sum Techniques.  
ICI Monograph No.3, Oliver & Boyd. Edinburgh.
- [75] WOLD, H. (1954). A Study in the Analysis of Stationary Time Series. Almqvist &  
Wiksell, Stockholm ( first edition 1938).
- [76] YOUNG, P. C. (1974). Recursive approaches to time series analysis. Bul. Inst.  
Maths. and Applications. 10. 209-224.
- [77] ZELLNER, A. (1971). An Introduction to Bayesian Inference in Economics. Wiley.